GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 09:34:06 ; Search time 1885.18 Seconds
(without alignments)
13009.846 Million cell updates/sec

Title: US-09-530-663B-5
Perfect score: 1172
Sequence: 1 tgtaggattggtggaatggg.....agcccgcaatggacgcgaca 1172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1172	100.0	1172	8	AF041051	AF041051 Populus t
2	826.8	70.5	4397	8	AY043494	AY043494 Populus t
c 3	83	7.1	104992	2	AC005504	AC005504 Plasmodiu
c 4	83	7.1	169546	2	AC004157	AC004157 Plasmodiu
c 5	82.2	7.0	93368	9	AL133402	AL133402 Human DNA
6	81.2	6.9	185699	2	AC021553	AC021553 Homo sapi
c 7	81	6.9	236120	14	AF063866	AF063866 Melanoplu
8	80.2	6.8	67970	3	PFMAL1P3	AL031746 Plasmodiu
c 9	80.2	6.8	202645	2	AL513330	AL513330 Homo sapi
10	78.8	6.7	181486	2	AC096550	AC096550 Homo sapi
c 11	78.4	6.7	14867	3	AE001398	AE001398 Plasmodiu
12	78.4	6.7	189214	2	AC074035	AC074035 Homo sapi
13	78.4	6.7	208319	2	AC108683	AC108683 Homo sapi
14	78	6.7	86827	3	PFMAL3P5	AL034556 Plasmodiu
c 15	78	6.7	192731	9	AC026698	AC026698 Homo sapi
c 16	77.8	6.6	172816	9	AC093899	AC093899 Homo sapi
17	77.6	6.6	102401	2	AC094148	AC094148 Rattus no
c 18	77.6	6.6	173480	9	AC090014	AC090014 Homo sapi
c 19	77.4	6.6	245802	2	AC006279	AC006279 Plasmodiu
20	77.2	6.6	104992	2	AC005504	AC005504 Plasmodiu
21	77.2	6.6	169546	2	AC004157	AC004157 Plasmodiu
22	76.8	6.6	158548	3	PFMAL3P2	AL034558 Plasmodiu
23	76.6	6.5	242159	2	AC091567	AC091567 Homo sapi
c 24	76.4	6.5	176506	2	AC103500	AC103500 Rattus no
c 25	76.2	6.5	53932	2	AC023371	AC023371 Homo sapi
c 26	75.8	6.5	116696	3	PFMAL3P3	Z98547 Plasmodium
c 27	75.8	6.5	186431	2	AC022281	AC022281 Homo sapi
28	75.6	6.5	196355	9	AC099557	AC099557 Homo sapi
29	75.4	6.4	17443	3	AC024806	AC024806 Caenorhab
c 30	75.2	6.4	34548	6	AX349036	AX349036 Sequence
31	75.2	6.4	178783	9	AC068139	AC068139 Homo sapi
c 32	75.2	6.4	205130	2	AC105425	AC105425 Homo sapi
33	75	6.4	44352	3	AF098501	AF098501 Caenorhab
34	75	6.4	152359	9	AC093790	AC093790 Homo sapi
35	75	6.4	182974	2	AC105148	AC105148 Homo sapi
36	74.8	6.4	969	3	DMA132902	AJ132902 Drosophil
c 37	74.8	6.4	85779	8	SCE011856	AJ011856 Saccharom
c 38	74.8	6.4	167471	9	AC019211	AC019211 Homo sapi
c 39	74.8	6.4	181792	9	AC098822	AC098822 Homo sapi
c 40	74.8	6.4	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 41	74.6	6.4	86827	3	PFMAL3P5	AL034556 Plasmodiu
c 42	74.6	6.4	88037	2	PFMAL13P8	AL096782 Plasmodiu
c 43	74.4	6.3	35077	3	CEY38H8A	AL021483 Caenorhab
c 44	74.4	6.3	118425	9	AL356005	AL356005 Human DNA
45	74.4	6.3	189790	2	AC107420	AC107420 Homo sapi

ALIGNMENTS

RESULT 1
AF041051
LOCUS AF041051 1172 bp DNA linear PLN 26-JUN-1998
DEFINITION Populus tremuloides clone Pt4CL1p 4-coumarate:CoA ligase gene, promoter region.
ACCESSION AF041051
VERSION AF041051.1 GI:3258638
KEYWORDS .
SOURCE quaking aspen.
ORGANISM Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 1172)
AUTHORS Hu,W.J., Kawaoka,A., Tsai,C.J., Lung,J., Osakabe,K., Ebinuma,H. and Chiang,V.L.
TITLE Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (Populus tremuloides)
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (9), 5407-5412 (1998)
MEDLINE 98226828
REFERENCE 2 (bases 1 to 1172)
AUTHORS Hu,W.-J. and Chiang,V.L.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1998) Plant Biotechnology Research Center, School
            of Forestry and Wood Products, Michigan Technological University,
            1400 Townsend Drive, Houghton, MI 49931, USA
FEATURES             Location/Qualifiers
     source          1. .1172
                     /organism="Populus tremuloides"
                     /db_xref="taxon:3693"
                     /clone="Pt4CL1p"
                     /tissue_type="leaf"
     gene            1. .>1172
                     /gene="4-coumarate:CoA ligase"
     promoter        1. .1172
                     /gene="4-coumarate:CoA ligase"
BASE COUNT      399 a    224 c    181 g    368 t
ORIGIN


  Query Match           100.0%;  Score 1172;  DB 8;  Length 1172;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-182;
  Matches 1172;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tgtaggattggtggaatgggatcattcctaatcccttaatgacggtggcatgaacacaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGTAGGATTGGTGGAATGGGATCATTCCTAATCCCTTAATGACGGTGGCATGAACACAAA 60

Qy      61 gcaaagagaagttaggtcactcctcctttatatatatatatatatgcatgcatgaggacc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCAAAGAGAAGTTAGGTCACTCCTCCTTTATATATATATATATATGCATGCATGAGGACC 120

Qy     121 atggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcactagtttt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGGCTATGATGAAGGTTAATAGAGGTAGTTGTGATTGAGATATGTCCAGCACTAGTTTT 180

Qy     181 ttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataatgaataata 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTGTTGGTGTGATTTCTCATGATGACGCGAAAATTTTATATATATATATAATGAATAATA 240

Qy     241 tgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgacc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGATTGATTATTCTCTGTAATTTTGTGAAATAGATTAAAACAGCTCAATGTGAGGTGACC 300

Qy     301 agttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaatttg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGTTGTCAAATGACCACTCGACTTGGGGCATGGTGATTTTTCAAATCACAACTCAATTTG 360

Qy     361 aaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggct 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAAACTAAAATTAAAAAAGATTTAGATTATTAAATTATTAGGTTAATTCACGGGTTGGCT 420

Qy     421 aatcaattattattaattaaaacgatagtatttttgataatttaattaaaattttattgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATCAATTATTATTAATTAAAACGATAGTATTTTTGATAATTTAATTAAAATTTTATTGG 480

Qy     481 atttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatat 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATTTGAATGAACTCAATTACATCACAAAAAACCTAATCAAATTAATATCTTATGTGATAT 540

Qy     541 aatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AATTTAGAAATATAAATGATTAACCTTTAAATCTCGAGTTTCTCTTATAAAAAACACGTA 600

Qy     601 taattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TAATTGGGCTAGATTTAACAGCTATTATTCAAACTGGCCAGGACAATTATTAAAATTAAT 660

Qy     661 aattattatttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AATTATTATTTTTTCTAATAAAGCACTTCCTAATTGTTAAAATATATGTCTAAACACTAA 720

Qy     721 taataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TAATAAAATTTATTTGTGTATCTTTGGCAGTAGGTGAGAGGTGCTGACAAATAAATTAGT 780

Qy     781 gcataaaatataatggattggtggtctgtgaaaagacaggtggaggacaagccacctctc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCATAAAATATAATGGATTGGTGGTCTGTGAAAAGACAGGTGGAGGACAAGCCACCTCTC 840

Qy     841 tcaagtcaaaaggccatttcacaaccaacccaaatgggaacccaccaccgttccccgcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCAAGTCAAAAGGCCATTTCACAACCAACCCAAATGGGAACCCACCACCGTTCCCCGCCA 900

Qy     901 ttaaaatccctaatctcaccaacccaactccacagattcttcaccaaacgcaactgattt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TTAAAATCCCTAATCTCACCAACCCAACTCCACAGATTCTTCACCAAACGCAACTGATTT 960

Qy     961 ttcaatcaatgttttccctatactaccccccccaacaactccataatacccaatttgtcct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCAATCAATGTTTTCCCTATACTACCCCCCCCAACAACTCCATAATACCCAATTTGTCCT 1020

Qy    1021 ttcaccaaccccccgtcctccgtgccagccaattctatatcagcaggaatgctctgcactc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTCACCAACCCCCGTCCTCCGTGCCAGCCAATTCTATATCAGCAGGAATGCTCTGCACTC 1080

Qy    1081 tgctttctcaggtctcctaccataagaaaacagagagcacctaaaactcgccatctctcc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TGCTTTCTCAGGTCTCCTACCATAAGAAAACAGAGAGCACCTAAAACTCGCCATCTCTCC 1140

Qy    1141 ctctgcatctttagcccgcaatggacgcgaca 1172
          ||||||||||||||||||||||||||||||||
Db    1141 CTCTGCATCTTTAGCCCGCAATGGACGCGACA 1172


RESULT   2
AY043494
LOCUS       AY043494                4397 bp    DNA     linear   PLN 17-SEP-2001
DEFINITION  Populus tomentosa 4-coumarate:CoA ligase gene, complete cds.
ACCESSION   AY043494
VERSION     AY043494.1  GI:15636676
KEYWORDS    .
SOURCE      Populus tomentosa.
  ORGANISM  Populus tomentosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 4397)
  AUTHORS   Lu,H., Zeng,Q. and Jiang,X.
  TITLE     Genomic DNA sequence of 4-coumarate:CoA ligase gene from Populus
            tomentosa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4397)
  AUTHORS   Lu,H., Zeng,Q. and Jiang,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2001) Biology, Beijing Forestry University,
            Beijing Qinghua East Road No. 35, Beijing 100081, China
FEATURES             Location/Qualifiers
     source          1. .4397
                     /organism="Populus tomentosa"
                     /db_xref="taxon:118781"
     promoter        1. .1136
     mRNA            join(1137. .2136,3321. .3519,3644. .3857,3957. .4059,
                     4174. .4397)
                     /product="4-coumarate:CoA ligase"
     CDS             join(1153. .2136,3321. .3519,3644. .3857,3957. .4059,
                     4174. .4284)
                     /codon_start=1
                     /product="4-coumarate:CoA ligase"
                     /protein_id="AAL02144.1"
                     /db_xref="GI:15636677"
                     /translation="MNPQEEFIFRSKLPDIYIPKNLPLHSYVLENLSNHSSKPCLING
                     ANGDVYTYADVELTARRVASGLNKIGIQQGDVIMLFLPSSPEFVLAFLGASHRGAIIT
                     AANPFSTPAELAKHAKASRAKLLITQACYYEKVKDFARESDVKVMCVDSAPDGCLHFS
                     ELTQADENEAPQVDISPDDVVALPYSSGTTGLPKGVMLTHKGLITSVAQQVDGDNPNL
                     YFHSEDVILCVLPMFHIYALNSIMLCGLRVGASILIMPKFEIGSLLGLIEKYKVSIAP
                     VVPPVMMSIAKSPDLDKHDLSSLRMIKSGGAPLGKELEDTVRAKFPQARLGQGYGMTE
                     AGPVLAMCLAFAKEPFDIKPGACGTVVRNAEMKIVDPETGASLPRNQPGEICIRGDQI
```

```
                   MKGYLNDPEATSRTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPAELE
                   ALLIAHPEISDAAVVGLKDEDAGEVPVAFVVKSEKSQATEDEIKQYISKQVIFYKRIK
                   RVFFIEAIPKAPSGKILRKNLKEKLAGI"
BASE COUNT     1394 a    823 c    819 g   1361 t
ORIGIN


  Query Match           70.5%;  Score 826.8;  DB 8;  Length 4397;
  Best Local Similarity   90.5%;  Pred. No. 6.8e-126;
  Matches 1035;  Conservative    0;  Mismatches   52;  Indels   57;  Gaps   12;

Qy      39 atgacggtggcatgaacacaaagcaaagagaagttaggtcactcctcctt---------- 88
           ||||||||||||||||||||||||||||||||||||||||||||||||||          
Db       1 ATGACGGTGGCATGAACACAAAGCAAAGAGAAGTTAGGTCACTCCTCCTTTTTATATATA 60

Qy      89 ----------tatatatatatatatatgcatgcatgaggaccatggctatgatgaaggtt 138
                    ||||||||||||||||  ||||||||||||||||||||||||||||||||  
Db      61 TATTTATATATATATATATATATATAGATATGCATGAGGACCATGGCTATGATGAAGGTT 120

Qy     139 aatagaggtagttgtgattgagatatgtccagcactagtttttttgttggtgtgatttctc 198
           |||||  |||||||||||| ||||||||||||||||||||||||||||||||||| ||||  
Db     121 AATAG-GGTAGTTGTGATAGAGATATGTCCAGCACTAGTTTTTTGTTGGTGTGATCTCTC 179

Qy     199 atgatgacgcgaaaattttata----------------tatatatataatgaataatatg 242
           |||||||| |||||||  ||||                ||||||||||   |   ||||  
Db     180 ATGATGACGGGAAAATTATATATATATATATATATATATATATATATATATATATATATA 239

Qy     243 attgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgaccag 302
           ||| |||||||||| |||||||||  ||||| ||| ||| || ||||||||||||||||||  
Db     240 ATTAATTATTCTCTCTAATTTTGTAAAATAAATTTAAA-AGGTCAATGTGAGGTGACCAG 298

Qy     303 ttgtcaaatgaccactcgactt-ggggcatggtgatttttcaaatcacaactcaatttga 361
           |||||||||||||||||||||| ||||||||  |||||||| |||||||||||||||||  
Db     299 TTGTCAAATGACCACTCGACTTGGGGGCATGATGATTTTTCGAATCACAACTCAATTTGA 358

Qy     362 aaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggcta 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db     359 AAACTAAAATTAAAAAAGATTTAGATTATTAAATTATTAGGTTAATTCACGGGTTGGCTA 418

Qy     422 atcaattattattaattaaaacgatagtatttttgataatttaattaaaatttttattgga 481
           ||||||||||  |||||||||  ||||||||||||||||||||||||||||||  ||||||  
Db     419 ATCAATTATTGTTAATTAAAATGATAGTATTTTTGATAATTTAATTAAAATTTAATTGGA 478

Qy     482 tttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatata 541
           ||||||||||  |||||||||  ||||||||||||||||||||||||||||||||||||  
Db     479 TTTGAATGAA-TCAATTACATTACAAAAAACCTAATCAAATTAATATCTTATGTGATATA 537

Qy     542 atttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtat 601
           | ||||||||||||||||||| ||||||||||||||||||||||||||||||  || |  
Db     538 A-TTAGAAATATAAATGATTAACATTTAAATCTCGAGTTTCTCTTATAAAAAACTTGTGT 596

Qy     602 aattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaata 661
           |||||||| ||||||||||||||||||||||||||||| ||||| ||||||||||||||  
Db     597 AATTGGGCTGGATTTAACAGCTATTATTCAAACTGGCCAAGACAACTATTAAAATTAATA 656

Qy     662 attattatttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaat 721
           ||||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db     657 ATTAAT-TTTTTTCTAATAAAGCACTTCCTAATTGTTAAAATATATGTCTAAACACTAAC 715

Qy     722 aataaaatttatttgtgtatctttggcagtaggtgag-----------aggtgctgacaa 770
           ||||||||||||||||||  || ||||||||||||          ||||||||||||  
Db     716 AATAAAATTTATTTGTGTATATTAGGCAGTAGGTGAGAGGAGAGACACAGGTGCTGACAA 775

Qy     771 ataaattagtgcataaaatataatggattggtggtctgtgaaaagacaggtggaggacaa 830
           ||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||  
Db     776 ATAAATTAGTGCATAAAATATAATGGATTGGTGGTCTGTGAAAAG-CAAGTGGAGGACAA 834

Qy     831 gccacctctctcaagtcaaaaggcca-tttcacaaccaacccaaatgggaacccaccacc 889
           ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||  
Db     835 GCCACCTCTCTCAAGTCAAAAGGCCACTTTCACAACCAACCCAAATGGGAACCCACCACC 894

Qy     890 gttccccgccattaaaatccctaatctcaccaacccaactccacagattcttcaccaaac 949
           |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||  
```

```
Db     895 GTTCCTCGCCATTAAAATCCCTAATCTCACCAACCCAACTCCACAGATTCTTCACCAAAC 954

Qy     950 gcaactgatttttcaatcaatgttttccctatactaccccccaacaactccataatacc 1009
           |||||||||||||||||||||||||||||||| ||||||||  |  |   |||||||||  
Db     955 GCAACTGATTTTTCAATCAATGTTTTCCCTATATTACCCCCC--AACAACTCATAATACC 1012

Qy    1010 caatttgtccttttcaccaaccccgtcctccgtgccagccaattctatatcagcaggaat 1069
           | ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| | ||  
Db    1013 CCATTTGTCCTTTCACCAACCCCCGTCCTCCGTGCCAGCCATTTCTATATCAGCATGGAT 1072

Qy    1070 gctctgcactctgctttctcaggtctcctaccataagaaaacagagagcacctaaaactc 1129
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||  
Db    1073 GCTCTGCACTCTGCTTTCTCAGGTCTCCTACCATTGAAAAACAGAGAGCACCTAAAACTC 1132

Qy    1130 gcca 1133
           ||||
Db    1133 GCCA 1136


RESULT   3
AC005504/c
LOCUS       AC005504               104992 bp    DNA     linear   HTG 01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC005504
VERSION     AC005504.3  GI:4558584
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337172.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    58642: contig of 58642 bp in length
            *    58643    58842: gap of unknown length
            *    58843    91011: contig of 32169 bp in length
            *    91012    91211: gap of unknown length
            *    91212   104992: contig of 13781 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104992
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT    44286 a   9326 c   9564 g  41411 t    405 others
ORIGIN


  Query Match           7.1%;  Score 83;  DB 2;  Length 104992;
  Best Local Similarity   45.3%;  Pred. No. 6.8e-05;
  Matches  302;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy     133 aaggttaatagaggtagttgtgattgagatatgtccagcactagtttttgttggtgtga 192
           ||   |||||    |  || |     | || |    |     | ||| || ||    | |  
Db   74707 AATAATAATATTTATTATTATATAATATATTTAATTAAATTAAATTTATTATTTAATTAA 74648

Qy     193 tttctcatgatgacgcgaaaattttatatatatatatatgaatgaatatatgattgattatt 252
```

```
Db  74647 TTTAAAATAAAATAAAAATAATAATATTTATTATTATATAATATATTTAATTAAATTAAA 74588

Qy    253 ctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgaccagttgtcaaatg 312
Db  74587 TTTATTATTTAATTAATTAATTTAAAATAAAATAAAAATAATAATATTTATTATTATATA 74528

Qy    313 accactcgacttggggcatggtgatttttcaaatcacaactcaatttgaaaactaaaatt 372
Db  74527 ATATATTTAATTAAATTAAATTTATTATTTAATTAATTTAAAATAAATAATAATTAAATT 74468

Qy    373 aaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaattatta 432
Db  74467 AATATATATTATAATTAATTAAAAATAAAATAATATTCAATTATAAATTTATTAATAATT 74408

Qy    433 ttaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaatgaac 492
Db  74407 TTAAATAAATAATTAAAATAATAATAAATTAAATAATTTAATTAATATAAATAAACATTA 74348

Qy    493 tcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataatttagaaata 552
Db  74347 TAAATTAAAAAATAATTTAATAAATATATATATAATTATTAATAAATTTAAATACATATA 74288

Qy    553 taaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgggctag 612
Db  74287 TTTTTAATAATAATTTCTTAATTTATTTTATTACATTATTATAATATATATTTTTTATTT 74228

Qy    613 atttaacagctattattcaaactggccaggacaattattaaaattaataattattatttt 672
Db  74227 AAAAAAATATAATTAAATAAATTAATTTAATAATTAAATATATTTAATAGTAATTAAATA 74168

Qy    673 ttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaaattta 732
Db  74167 TTAAACAAATAATATAAATATTATATAATATTATTAATTAAATTAAAATAATAATTTATT 74108

Qy    733 tttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaaatata 792
Db  74107 TTATAATTATATATATATATTAATAATTAATTTAAAATATAAATTAAGAAAGATAATTTT 74048

Qy    793 atggatt 799
Db  74047 ATACTTT 74041


RESULT  4
AC004157/c
LOCUS       AC004157              169546 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC004157
VERSION     AC004157.8  GI:9797712
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 169546)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169546)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810447.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23466: contig of 23466 bp in length
            *   23467    23666: gap of unknown length
            *   23667   169546: contig of 145880 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169546
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC293"
                     /clone="3D7"
BASE COUNT    69871 a  15381 c  15705 g  68389 t    200 others
ORIGIN


  Query Match             7.1%;  Score 83;  DB 2;  Length 169546;
  Best Local Similarity   45.3%;  Pred. No. 6.1e-05;
  Matches  302;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy    133 aaggttaatagaggtagttgtgattgagatatgtccagcactagttttttgttggtgtga 192
Db  86291 AATAATAATATTTATTATTATATAATATATTTAATTAAATTAAATTTATTATTTAATTAA 86232

Qy    193 tttctcatgatgacgcgaaaatttatatatatatataatgaataatatgattgattatt 252
Db  86231 TTTAAAATAAAATAAAAATAATAATATTTATTATTATATAATATATTTAATTAAATTAAA 86172

Qy    253 ctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgaccagttgtcaaatg 312
Db  86171 TTTATTATTTAATTAATTAATTTAAAATAAAATAAAAATAATAATATTTATTATTATATA 86112

Qy    313 accactcgacttggggcatggtgatttttcaaatcacaactcaatttgaaaactaaaatt 372
Db  86111 ATATATTTAATTAAATTAAATTTATTATTTAATTAATTTAAAATAAATAATAATTAAATT 86052

Qy    373 aaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaattatta 432
Db  86051 AATATATATTATAATTAATTAAAAATAAAATAATATTCAATTATAAATTTATTAATAATT 85992

Qy    433 ttaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaatgaac 492
Db  85991 TTAAATAAATAATTAAAATAATAATAAATTAAATAATTTAATTAATATAAATAAACATTA 85932

Qy    493 tcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataatttagaaata 552
Db  85931 TAAATTAAAAAATAATTTAATAAATATATATATAATTATTAATAAATTTAAATACATATA 85872

Qy    553 taaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgggctag 612
Db  85871 TTTTTAATAATAATTTCTTAATTTATTTTATTACATTATTATAATATATATTTTTTATTT 85812

Qy    613 atttaacagctattattcaaactggccaggacaattattaaaattaataattattatttt 672
Db  85811 AAAAAAATATAATTAAATAAATTAATTTAATAATTAAATATATTTAATAGTAATTAAATA 85752

Qy    673 ttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaaattta 732
Db  85751 TTAAACAAATAATATAAATATTATATAATATTATTAATTAAATTAAAATAATAATTTATT 85692

Qy    733 tttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaaatata 792
Db  85691 TTATAATTATATATATATATTAATAATTAATTTAAAATATAAATTAAGAAAGATAATTTT 85632

Qy    793 atggatt 799
Db  85631 ATACTTT 85625


RESULT  5
AL133402/c
LOCUS       AL133402              93368 bp    DNA     linear   PRI 24-JUL-2000
DEFINITION  Human DNA sequence from clone RP5-1077H22 on chromosome 6. Contains
```

```
          a putative novel gene, ESTs, STSs and GSSs, complete sequence.
ACCESSION AL133402
VERSION   AL133402.10  GI:8039186
KEYWORDS  HTG.
SOURCE    human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 93368)
  AUTHORS   Milne,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On May 23, 2000 this sequence version replaced gi:7939106.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP5-1077H22 is from the library RPCI-5 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP5-1077H22.
FEATURES             Location/Qualifiers
     source          1. .93368
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP5-1077H22"
                     /clone_lib="RPCI-5"
     repeat_region   231. .537
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   638. .892
                     /note="AluSx repeat: matches 39. .303 of consensus"
     repeat_region   1142. .1329
                     /note="MIR repeat: matches 30. .235 of consensus"
     repeat_region   1386. .1680
                     /note="AluSq repeat: matches 1. .304 of consensus"
     misc_feature    complement(3971. .4298)
                     /note="match: GSS: Em:AQ720689"
     repeat_region   4299. .4722
                     /note="L1PBa repeat: matches -1546. .-1136 of consensus"
     repeat_region   4721. .6631
                     /note="L1PBa repeat: matches -134. .2762 of consensus"
     repeat_region   6632. .6939
                     /note="AluY repeat: matches 3. .310 of consensus"
     repeat_region   6940. .7827
                     /note="L1PBa repeat: matches 2762. .3638 of consensus"
     repeat_region   7856. .9561
                     /note="L1PB2 repeat: matches 4121. .5791 of consensus"
     misc_feature    9565. .9598
                     /note="match: GSS: Em:AQ984702"
     repeat_region   9599. .9750
                     /note="4 copies 38 mer 71% conserved"
     repeat_region   9604. .9739
                     /note="68 copies 2 mer ta 73% conserved"
     misc_feature    9708. .9831
                     /note="match: STS: Em:G34561"
     misc_feature    complement(9730. .9849)
                     /note="match: STS: Em:Z52156"
     misc_feature    9764. .9824
                     /note="match: GSS: Em:AQ014122 Em:AQ598220"
     misc_feature    complement(9764. .9823)
                     /note="match: GSS: Em:AG019010 Em:AG026881"
     misc_feature    complement(9764. .9824)
                     /note="match: GSS: Em:AQ933387 Em:AZ023541"
     misc_feature    9772. .9824
                     /note="match: GSS: Em:AZ110479"
     misc_feature    complement(9796. .9835)
                     /note="match: GSS: Em:AZ002912"
     misc_feature    9796. .9835
                     /note="match: GSS: Em:AL214116 Em:AQ341248"
     misc_feature    9796. .9836
                     /note="match: GSS: Em:AQ548435"
     misc_feature    complement(9796. .9828)
                     /note="match: GSS: Em:AQ324044 Em:AQ373503"
     misc_feature    complement(9796. .9824)
                     /note="match: GSS: Em:AQ598220"
     misc_feature    9796. .9823
                     /note="match: GSS: Em:AQ745266"
     misc_feature    9796. .9824
                     /note="match: GSS: Em:AQ014122"
     misc_feature    9797. .9828
                     /note="match: GSS: Em:AQ745589"
     misc_feature    9797. .9824
                     /note="match: GSS: Em:AQ014122"
     misc_feature    9797. .9841
                     /note="match: GSS: Em:AQ745266"
     misc_feature    9803. .9824
                     /note="match: GSS: Em:AQ598220"
     misc_feature    9806. .9841
                     /note="match: GSS: Em:AG026544"
     misc_feature    9807. .9841
                     /note="match: GSS: Em:AZ028723"
     repeat_region   9842. .10202
                     /note="L1PB2 repeat: matches 5782. .6154 of consensus"
     repeat_region   10724. .11003
                     /note="AluJo repeat: matches 1. .289 of consensus"
     repeat_region   11341. .11501
                     /note="MER5B repeat: matches 5. .173 of consensus"
     repeat_region   12016. .12384
                     /note="MLT1B repeat: matches 1. .387 of consensus"
     repeat_region   13221. .13307
                     /note="HAL1 repeat: matches 550. .638 of consensus"
     repeat_region   13294. .13693
                     /note="HAL1 repeat: matches 595. .1036 of consensus"
     repeat_region   14165. .14463
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   14897. .15051
                     /note="L1M4c repeat: matches 267. .420 of consensus"
     repeat_region   15208. .15761
                     /note="MER66B repeat: matches 1. .486 of consensus"
     repeat_region   17062. .17326
                     /note="AluJb repeat: matches 28. .294 of consensus"
     misc_feature    complement(19549. .20029)
                     /note="match: GSS: Em:AQ734906"
     mRNA            join(<21618. .21754,25301. .>25566)
                     /gene="dJ1077H22.1"
                     /note="match: ESTs: Em:AA960797"
                     /evidence=not_experimental
                     /product="dJ1077H22.1 (putative novel protein)"
     gene            join(21618. .21754,25301. .25566)
                     /gene="dJ1077H22.1"
     repeat_region   24053. .24358
                     /note="MER33 repeat: matches 7. .323 of consensus"
     repeat_region   25456. .25647
                     /note="MIR repeat: matches 43. .246 of consensus"
     repeat_region   26551. .26628
```

```
                /note="MIR repeat: matches 44. .121 of consensus"
     repeat_region  27163. .27450
                /note="AluSc repeat: matches 1. .288 of consensus"
     repeat_region  28099. .28569
                /note="MLT1D repeat: matches 2. .505 of consensus"
     repeat_region  32415. .32450
                /note="L2 repeat: matches 2677. .2710 of consensus"
     repeat_region  34405. .34634
                /note="L1MB8 repeat: matches 5857. .6120 of consensus"
     repeat_region  34944. .36430
                /note="L2 repeat: matches 62. .1680 of consensus"
     repeat_region  36549. .36873
                /note="L2 repeat: matches 2411. .2750 of consensus"
     repeat_region  37369. .37635
                /note="MER33 repeat: matches 22. .294 of consensus"
     repeat_region  37654. .37931
                /note="L1MC5 repeat: matches 7507. .7777 of consensus"
     repeat_region  38000. .38111
                /note="28 copies 4 mer tata 67% conserved"
     repeat_region  38001. .38112
                /note="56 copies 2 mer at 67% conserved"
     repeat_region  39562. .39844
                /note="AluY repeat: matches 1. .281 of consensus"
     repeat_region  41294. .41412
                /note="MIR repeat: matches 53. .171 of consensus"
     repeat_region  41413. .41819
                /note="MSTB repeat: matches 1. .426 of consensus"
     repeat_region  41820. .41874
                /note="MIR repeat: matches 171. .225 of consensus"
     repeat_region  42349. .42490
                /note="MIR repeat: matches 34. .185 of consensus"
     misc_feature   43137. .43480
                /note="match: GSS: Em:AQ430150"
     repeat_region  44406. .44524
                /note="MIR repeat: matches 21. .146 of consensus"
     repeat_region  45108. .45474
                /note="MER74A repeat: matches 132. .528 of consensus"
     repeat_region  45548. .45644
                /note="MIR repeat: matches 56. .147 of consensus"
     repeat_region  46184. .46239
                /note="14 copies 4 mer agaa 73% conserved"
     repeat_region  46186. .46375
                /note="5 copies 38 mer 65% conserved"
     repeat_region  46228. .46383
                /note="4 copies 39 mer 76% conserved"
     repeat_region  46452. .46565
                /note="3 copies 38 mer 75% conserved"
     repeat_region  46474. .46565
                /note="23 copies 4 mer ggag 65% conserved"
     repeat_region  46483. .46558
                /note="38 copies 2 mer gg 67% conserved"
     repeat_region  47326. .47676

  Query Match          7.0%;  Score 82.2;  DB 9;  Length 93368;
  Best Local Similarity  47.7%;  Pred. No. 9.5e-05;
  Matches  277;  Conservative    0;  Mismatches  298;  Indels    6;  Gaps    1;

Qy     210 aaaatttttatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
           | |  ||||||  |||||||  ||  ||||  ||    | |||  |       ||   ||  |  |
Db   77802 ATACATTTATAATTATATATTATATATAAAATATATAATTGTATATAATATAATTATATA 77743

Qy     270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
           ||  ||  |||      |||   |  | |     |   | || |    |  | |
Db   77742 TTATATATAAAATATATAATTGTATATAATATAATTATATATTATATATATAATATATAA 77683

Qy     330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
           ||   ||  |  || |    | |  || |  ||  || ||||  | |  || ||  |
Db   77682 TTGTATATAATATAATTATATATTATATATATAATATATAATTGTATATAATATAATTAT 77623

Qy     390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
           || || |     |||||       |   ||||  | |  |||  ||  |||  |||
Db   77622 ATATATAAAATATATAATT------GTATATAATATAATTATATATATATAAAATATATA 77569

Qy     450 atttttgataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaa 509
           ||| |  ||||| ||||||| |  ||   |  |    |   || |  ||   ||   | |
Db   77568 ATTGTATATAATATAATTATATATTATATATAAAATATATAATTGTATATAATATAATTA 77509

Qy     510 aacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta 569
           |  | ||  |   ||||| | ||   |||||||  ||   |||||    | | |  | ||
Db   77508 TATATTATATATAAAATATATAATTGTATATAATATAATTATATATTATAATTAAATATA 77449

Qy     570 aatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattatt 629
           |  |  |  |  | | || | | |     || |     ||  | ||    ||| |||
Db   77448 TAATTGTATATAATATAATTATATATTATAATTAAATATATAATTGTATATAATATAATT 77389

Qy     630 caaactggccaggacaattattaaaattaataattattatttttttctaataaagcacttc 689
           |  |       | |||   |||   ||   | ||| ||| | | | ||| |  |  |
Db   77388 ATATATTATATATAAAATATATAATTGTATATAATATAATTATATATTATATATAAAATA 77329

Qy     690 ctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggca 749
           | |  || |  ||||   || | | ||   ||||||| |||   |||||  |     |
Db   77328 TATAATTGTATATAATATAATTATATATTATATATAAAATATATAATTGTATATAATATA 77269

Qy     750 gtaggtgagaggtgctgacaaataaattagtgcataaaata 790
           |   |       | | | ||| | | ||  |||| |||
Db   77268 ATTATATATTATATATAAAATATATAATTGTATATAATATA 77228


RESULT   6
AC021553
LOCUS       AC021553              185699 bp    DNA     linear   HTG 05-JAN-2002
DEFINITION  Homo sapiens chromosome 15 clone RP11-709B3 map 15, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC021553
VERSION     AC021553.13  GI:15706183
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185699)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-709B3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185699)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 20, 2001 this sequence version replaced gi:15148276.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5662
```

```
           Center clone name: 709_B_3
          ---------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1   155157: contig of 155157 bp in length
          *  155158 155257: gap of      100 bp
          *  155258   185699: contig of 30442 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185699
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-709B3"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    49597 a  42163 c  43800 g  50039 t    100 others
ORIGIN


  Query Match             6.9%;  Score 81.2;  DB 2;  Length 185699;
  Best Local Similarity   39.9%;  Pred. No. 0.00012;
  Matches  224;  Conservative    0;  Mismatches  338;  Indels    0;  Gaps    0;

Qy     174 tagttttttgttggtgtgatttctcatgatgacgcgaaaattttatatatatatataatg 233
           || | | |   |   |   ||||   |  ||      ||||| ||| ||  | |  |
Db  155086 TATTATGTATTATATTATATTTTATAAAATATTATAAAAATATTAAATTAAAAATTTTAA 155145

Qy     234 aataatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtga 293
           ||||||||  ||
Db  155146 AATAATATTATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 155205

Qy     294 ggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaact 353
                                                                |  | |
Db  155206 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATTATT 155265

Qy     354 caatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgg 413
           | | |  ||  ||  ||  |  |    ||| || ||  | |   ||   |   | |
Db  155266 TATTATATAATATATTATATATTATTTATTATATAATATATTATATATTATTTATTATAT 155325

Qy     414 gttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatt 473
             |   | ||  ||||||  | ||  ||   ||  ||| ||   || | ||  |||| |
Db  155326 AATATATTATATATTATTTATTATATAATATATTATATATTATTTATTATATATAAATAT 155385

Qy     474 ttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttat 533
           | ||   ||   ||   | | || || |  |  |  ||  |  | ||| || |||   ||
Db  155386 TAATATAATAATAAATTAATAAAATATAAAATTATTAATATTTTAAAAATATTATAAAAT 155445

Qy     534 gtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaa 593
           |  || |||  |  | | ||||  ||         | | |  | | | |   |||  |
Db  155446 ATTTTAAAATAAAATATTTTAAAATATATTTAAAATATTATAAAATATTTAAAATATTAT 155505

Qy     594 acacgtataattgggctagatttaacagctattattcaaactggccaggacaattattaa 653
           |    |  |||     | ||||||  |  |  |||| ||| |      |  |||   ||
Db  155506 ATTTAAAATATTATAAAATATTTAAAATATTATATTTAAAATATTATAAAATATTTAAAA 155565

Qy     654 aattaataattattatttttctaataaagcacttcctaattgttaaaatatatgtctaa 713
           ||||    | |    || |  | |  |   | ||   |  | ||||||||||| |   |
Db  155566 TATTATATTTAAAATATTATATTTAATATATATTTAAAATATTTTAAAATATATATTATA 155625

Qy     714 acactaataataaaatttattt 735
           | |    ||| | | | |||||
Db  155626 AAATATTTAAAATATTATATTT 155647


RESULT   7
AF063866/c
```

```
LOCUS       AF063866               236120 bp    DNA     linear   VRL 22-DEC-1998
DEFINITION  Melanoplus sanguinipes entomopoxvirus, complete genome.
ACCESSION   AF063866
VERSION     AF063866.1  GI:4049647
KEYWORDS    .
SOURCE      Melanoplus sanguinipes entomopoxvirus.
  ORGANISM  Melanoplus sanguinipes entomopoxvirus
            Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
            Entomopoxvirus B.
REFERENCE   1  (bases 1 to 236120)
  AUTHORS   Afonso,C.L., Tulman,E.R., Lu,Z., Oma,E., Kutish,G.F. and Rock,D.L.
  TITLE     The genome of Melanoplus sanguinipes entomopoxvirus
  JOURNAL   J. Virol. 73 (1), 533-552 (1999)
  MEDLINE   99102612
REFERENCE   2  (bases 1 to 236120)
  AUTHORS   Afonso,C.L., Tulman,E.R., Lu,Z., Oma,E., Kutish,G.F. and Rock,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Microbiology, Plum Island Animal Disease
            Center, U.S. Dept. Agriculture, Agricultural Research Service, P.O.
            Box 848, Greenport, NY 11944-0848, USA
FEATURES             Location/Qualifiers
     source          1. .236120
                     /organism="Melanoplus sanguinipes entomopoxvirus"
                     /isolate="Tucson"
                     /db_xref="taxon:83191"
     repeat_region   1. .7201
                     /note="left inverted terminal repeat"
                     /rpt_type=inverted
                     /evidence=experimental
     gene            complement(2443. .2631)
                     /gene="MSV001"
     CDS             complement(2443. .2631)
                     /gene="MSV001"
                     /codon_start=1
                     /product="ORF MSV001 62 amino acid repeat gene family
                     protein"
                     /protein_id="AAC97866.1"
                     /db_xref="GI:4049906"
                     /translation="MFLYLISQNINNVSKISISNTQSMFLYLISQNINNVSKISISNT
                     QSMFLYLISQNINNVSKI"
     gene            complement(6051. .6320)
                     /gene="MSV002"
     CDS             complement(6051. .6320)
                     /gene="MSV002"
                     /codon_start=1
                     /product="ORF MSV002 89 amino acid repeat gene family
                     protein"
                     /protein_id="AAC97867.1"
                     /db_xref="GI:4049907"
                     /translation="MAEKRNVKVKLVEPDVNKSLSTIIEALQNEIQKSIQDQSNILKQ
                     NLSEDIMNEVEKFESQVTELTNMLKNYVISTIEEKIDNYVNSLIS"
     gene            complement(6347. .6814)
                     /gene="MSV004"
     CDS             complement(6347. .6814)
                     /gene="MSV004"
                     /codon_start=1
                     /product="ORF MSV004 155 amino acid repeat gene family
                     protein"
                     /protein_id="AAC97868.1"
                     /db_xref="GI:4049908"
                     /translation="MYSCKMEIDLECVVLCKNCNVSSYQYTDGICINCDFPINIKLDI
                     CNKCNKINFVHDGVCGECYANKIVCNGDNTKSTCYIIIINDQKVIEEGVVESSKDIFI
                     SDKFTYNIINVVIISNNKILQINNIDKKYEKVFNRYLIFNYDHVILTQQIDEI"
     gene            complement(6574. .6756)
                     /gene="MSV003"
     CDS             complement(6574. .6756)
                     /gene="MSV003"
                     /codon_start=1
                     /product="ORF MSV003 60 amino acid repeat gene family
                     protein"
                     /protein_id="AAC97869.1"
                     /db_xref="GI:4049909"
                     /translation="MYHLINIQMVYALIVIFLLILNLIYVINVTKLILFMMEYAVNVM
```

```
                LTKLYVMVIIQNLHVI"
gene            complement(7431. .8171)
                /gene="MSV005"
CDS             complement(7431. .8171)
                /gene="MSV005"
                /codon_start=1
                /product="ORF MSV005 hypothetical protein"
                /protein_id="AAC97860.1"
                /db_xref="GI:4049900"
                /translation="MKMKILFKDYNTKIIYEYIKSLDCGYNLKKKLNIDDKLSKSIIK
                LTCYNPNISECVLPISIIKFIYTLSEIIILILLLNYYIKMQLNIKRAKHNMNYKKKYC
                KECELITFQNQFEICEICYKNKVFMYCSHCTDITEVDELNGKCKNCNFINKSYYAYIG
                VLMYKKYRIKCIICGNYDYISNICSEECKYKYLFKRLFTKRCIQCNMFIDSELNTKFC
                HYCYFYYYCYWYGYYWYIYPLIFDNYNG"
gene            8831. .9058
                /gene="MSV006"
CDS             8831. .9058
                /gene="MSV006"
                /codon_start=1
                /product="ORF MSV006 hypothetical protein"
                /protein_id="AAC97608.1"
                /db_xref="GI:4049648"
                /translation="MFESLFSDMCNILRLINDSKNVKFLILQEIHFNSINDISNVHSY
                IFEFLLPDTSNILILINGSKNLKSIISQSLQ"
gene            complement(9825. .11327)
                /gene="MSV008"
CDS             complement(9825. .11327)
                /gene="MSV008"
                /codon_start=1
                /product="ORF MSV008 leucine rich repeat gene family
                protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
                SW:P28854"
                /protein_id="AAC97859.1"
                /db_xref="GI:4049899"
                /translation="MNNINLYCYKNILKYLDNDDKNNLGIIYELFKYYNCILKVPIYY
                KYFTYYNLEYIKNLNCSDCNITDFKFLESLINLEILDISYNKNSNIYKCKLSSLLKNL
                NCNECKIIDFKFLESLINLEILDISYNKNSNIYKCKLSSLLKNLNCNECKIIDFKFLE
                SLINLEILDVSYNKNSNIYECTLPISLIELKCISCKIIDFKFLESLINLEILDVSSNE
                YSNISDCKLPVSLIKLNCEFCYIYDFKFLETLNNIKILNISYNKNSNISDCKKLPVSL
                IKLNCGYCYIYDFKFLEPLINLQKLNISSNSYSNISKCKLPISLIKFKSEYCDITNFK
                FLERLINLETLTIYMNSRSNISICKLPTSLINLNFIEYDIKNIKFIEKLINLKSLALS
                SRSSRNISQCILPTSLIVLDLYSCHIKNFKFLEPLINLKKLNISDNGDSNISKCKLPI
                SLIELYCSKCNIKNFNFLQPLINLQKLDISVNDFYSFNILNHFNYPPFIN"
gene            10282. .10476
                /gene="MSV007"
CDS             10282. .10476
                /gene="MSV007"
                /codon_start=1
                /product="ORF MSV007 hypothetical protein"
                /protein_id="AAC97609.1"
                /db_xref="GI:4049649"
                /translation="MYIVNVSKLINRSKNLKFVISQYSDLNLINDIGNLHFDIFEYEL
                ELIFNFCKLINGSKNLKSYI"
gene            complement(11374. .12138)
                /gene="MSV009"
CDS             complement(11374. .12138)
                /gene="MSV009"
                /codon_start=1
                /product="ORF MSV009 leucine rich repeat gene family
                protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
                SW:P28854"
                /protein_id="AAC97858.1"
                /db_xref="GI:4049898"
                /translation="MNNLTYYKEKLKKLNNDLNKIHKLAARNKLYRKYNSILNIPLHE
                NHIKYYNPKHVRKLQCSGKIPNNLEKLINLECLTICCASKIESKLPISLKKIVFFNCK
                ITDFKFLENLYNLEKLFIINNKKSNISDCKLPISIIEFCCNKCEITDFKFLEDLYNLK
                KLYISGNKNSNISHCNLPIKLTELECSYCNILDFKFLKNLYNLKKLFITGNNNSNISN
                CELPFKLIKFVCINCNITDFNFLKSSNNLRIITISN"
gene            complement(12364. .14199)
                /gene="MSV010"
CDS             complement(12364. .14199)
                /gene="MSV010"
                /codon_start=1
                /product="ORF MSV010 leucine rich repeat gene family
                protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
                SW:P28854"
                /protein_id="AAC97857.1"
                /db_xref="GI:4049897"
                /translation="MNNINLYCYENLIKYLKDSDIHNLASICELFQYYNYQLNICVNH
                YYKNYYKPKYIRNFNSSDRFGLTITDFTFLEELNNLEILDISNNEKLNISKCKLPNKI
                KKFICVRCDITDFKFLEPLINLEVLDISYNINSNISNYKLSKQLKEFICEKCNITDFT
                FLKKLNNLKVLDISENYISNISKCKLPKTIKKFICDRCDITDFTFLEELINLEILDIS
                NNIISNISNYKLSKTIKKFICARCAITDFTFLEELINLEILDVSYNHKLNISECELPI
                SLKKLYCNNCFIKNNTLLKKLINLTILNISFNKITDFKFLENLTNLEILDISENKNSN
                ISKCKLSKTIKKFICSRCAITDFKFLEHLTNLEILDVSYNYEANISEFKLSISLKELN
                CSDCYKTKDFKFLERLINLQKLNISYTFSSNISTCELTSTLVELNCSTCKINNFTFIE
                KLYNLKILDINNNPNSNISECKLSTALIELDCTICNITDFKFLEPLINLQKLNICGNH
                DSNISICKLPLSLIELNCMYCYISDFKFLEPLKSLKILNICKNGSSNIIYCNLSPTLE
                TIFYAKHDIDFEYINNSLHKLVYDFTSDSDSDSDSDSDSDSGIYD"
gene            14293. .15810
                /gene="MSV011"
CDS             14293. .15810
                /gene="MSV011"
                /codon_start=1
                /product="ORF MSV011 leucine rich repeat gene family
                protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
                SW:P28854"
                /protein_id="AAC97610.1"
                /db_xref="GI:4049650"
                /translation="MNNINLYCYENLLKYLDDKNKHKLAKVSKIFQYYNYQLNIPIHY
                SYFSHYKSKYIKILNCEKCNIKDFKFLEKLSNLEILNISDNKNSNISKCNLPNKIKKF
                ICVRCDIKDFNFLERLFRLQMLDISYNYIKSNISKCKLPISLIELYCKNCTNNNFYFL
                NDLPRLEILDISCNNKIININNIRLSKKLIKLNCSNNNISDIKFLEHLSNLEILNISY
                NKISYIYNYKLPNNLIELNCKYNNIKNFKFIERLHNLKKLNISNNKKSNISKCKLPQT
                IINLNIKACDITDFSFLEPLNNLEKLNITEENYDNETNLSKCKLPVSLRVLKSTDFNV
                CNYLFLERLINLEILYINDFFDFESILIRSLPISLKEINLSSCCYIDDIKFLEPLINL
                KKLNISDNRNLKITECKLPISLIELDCSSCDITDFKFLEPLINLKKLNISDSRNLKII
                ECKLPISLIELDCSLCDINDFKFLEPLTNLKTLTISFNTNILECNLPSNLKIINK"
gene            15855. .16247
                /gene="MSV012"
CDS             15855. .16247
                /gene="MSV012"
                /codon_start=1
                /product="ORF MSV012 hypothetical protein"
                /protein_id="AAC97611.1"
                /db_xref="GI:4049651"
                /translation="MNNINLYCYEISNICKFFSFSKNLKLVILELLHLNSINLFGNLH
                SDIFEFSLQNIFNIFRLHKDSKNLKSIISLIIQSNSINDIGNLQLDIFEFLLYDIINF
                FKLCNFSKNVKSIILQDEQSNILIDEGI"
gene            complement(16436. .16654)
                /gene="MSV013"
CDS             complement(16436. .16654)
                /gene="MSV013"
                /codon_start=1
                /product="ORF MSV013 leucine rich repeat gene family
                protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
                SW:P28854"

  Query Match             6.9%;  Score 81;  DB 14;  Length 236120;
  Best Local Similarity   46.6%;  Pred. No. 0.00012;
  Matches  296;  Conservative    0;  Mismatches  335;  Indels    4;  Gaps    1;

Qy    173 ctagttttttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataat 232
          ||| || |||     | |      | || || |    |  ||||| |||||      ||
Db 228394 CTAATTCTTTCAAATTATCTAAATTAATTATAAAATTAGTATTTTTTATATTACATGAAC 228335

Qy    233 gaataatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtg 292
           |||| |    || |||| | | | |  |  ||   ||    ||| |  |      |  |
Db 228334 TAATATTTAATTTAATTAATGTTTTTGGTAATGATGAAATATTTATATTATAATTGTTAG 228275

Qy    293 aggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaac 352
          |  |  ||| || |   |     | |  | ||    |     |   || | |    | ||
Db 228274 ATATATCCAATTCTTTTAAATTAATTAAATTTTTTAAAGAACTAGCATTTATTATATTAC 228215

Qy    353 tcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacg 412
              | | |  |||  | |  || ||||    || ||||||||||||   |   |  |
```

```
Db 228214 AATTACTAATATTTAATTTAATTAACGATT----TTGTTAAATTATTAATATCTATTTCA 228159

Qy    413 ggttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaat 472
          |    |   || ||| ||| |||| |   ||   ||| ||   |||| | | ||  |
Db 228158 TTTGAAAAATCTAAATATAATTTATTAGATATATCTAATTCTTTTAAATTAATTAAATTT 228099

Qy    473 tttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatctta 532
          |||| | ||| | ||  | |  ||| ||      || |  |||| |||||||   ||
Db 228098 TTTAAAGAATTAGCATTTATTATATTGCAATTACTAATATTTAATTTAATTAATGATTTT 228039

Qy    533 tgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaa 592
          || |  || |         |  ||    ||  | ||||||     | |    ||||
Db 228038 GGTAAATTATTAATATCTATTTCATTTGGAAAATCTAAATCCAATTTATTAGATATATTT 227979

Qy    593 aacacgtataattgggctagatttaacagctattattcaaactggccaggacaattatta 652
          ||  | | ||| |    || ||||   || |||   |||   |    |  ||| ||| ||
Db 227978 AATTCCTGTAAATTATTTAAATTTTTGAGGTATCTATCATTTTCAATATAACATTTACTA 227919

Qy    653 aaattaataattattatttttttctaataaagcacttcctaattgttaaaatatatgtcta 712
          | ||| |   | |||| |  |||| ||| |    |   ||   ||     |||
Db 227918 ATATTTAATTTAATTAATGATTCTGGTAATTCGTAGCACCATATGTATTTATTATTAGAT 227859

Qy    713 aacactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaat 772
          |   |||||| |   |  |  ||     |  ||  |      |||  |  |   | ||
Db 227858 ATATCTAATATTTGTAAATTATTAAAATTTCTTAAAAAATTAAAAGATTTAATATTACAT 227799

Qy    773 aaattagtgcataaaatataatggattggtggtct 807
          || |      |||| ||       |||||  | |
Db 227798 GAACTGCAATTTAAATTAATTAAATTTGGTAATTT 227764


RESULT   8
PFMAL1P3
LOCUS       PFMAL1P3               67970 bp    DNA     linear   INV 15-DEC-1999
DEFINITION  Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION   AL031746
VERSION     AL031746.9  GI:6594243
KEYWORDS    HTG.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 67970)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 16, 1999 this sequence version replaced gi:5763807.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
FEATURES             Location/Qualifiers
     source          1. .67970
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="1"
     gene            complement(join(1748. .2598,2748. .2848,2990. .3276))
                     /gene="MAL1P3.01"
     CDS             complement(join(1748. .2598,2748. .2848,2990. .3276))
                     /gene="MAL1P3.01"
                     /note="MAL1P3.01, conserved hypothetical protein, len: 412
                     aa, similarity: UPF0006 family eg to
                     YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
                     scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
                     overlap)"
                     /codon_start=1
                     /product="conserved hypothetical protein, UPF0006 family"
                     /protein_id="CAB63556.1"
                     /db_xref="GI:6594244"
                     /translation="MKLVFHYIKYINVLFYISIIFLKSNSLKIYNDLRYISTVNKYKV
                     LQIKKRSNLKKNHNIRKMEDNESSFIDIGSNLTDKMFDGVYNSKKHENDLQNVLNRAK
                     NNNVDKIIITCTCLAEIDKSLKICETYDPEGKFLYLSAGVHPTNCYEFIDKNKHEEKE
                     IIAKKEYEEFIKYFKNEQVENSKMENGNKKICDGEKDMNNLNEILLEKNLDTIPGFKY
                     NEKDKEYLENLKNKIIKYPNRIVCIGEIGLDFDRLYFCSKYIQIKYFIFQLKLVQMFN
                     LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVIHSFTDKEDIVHIIVQNYKNLYIGVNG
                     CSLKSLENINAVKKIPLNLLLLETDAPWCGVKKTHASYEYIKDTYEKRAYTNLKKIKN
                     IIKCDDNTIFKERNEPYNIA"
misc_feature         complement(2599. .2610)
                     /gene="MAL1P3.01"
                     /note="potential splice acceptor sequence"
misc_feature         complement(2742. .2747)
                     /gene="MAL1P3.01"
                     /note="potential splice donor sequence, atg/gttaaa"
misc_feature         complement(2849. .2861)
                     /gene="MAL1P3.01"
                     /note="potential splice acceptor sequence"
misc_feature         complement(2984. .2989)
                     /gene="MAL1P3.01"
                     /note="potential splice donor sequence, aaa/gtaaaa"
gene                 5005. .5496
                     /gene="MAL1P3.02"
CDS                  5005. .5496
                     /gene="MAL1P3.02"
                     /note="MAL1P3.02, hypothetical protein, len: 163 aa,
                     contains possible signal sequence"
                     /codon_start=1
                     /product="hypothetical protein, MAL1P3.02"
                     /protein_id="CAB63557.1"
                     /db_xref="GI:6594245"
                     /translation="MKLLNNRFVVLCPIIILFFFLNSVVLGNNNRNNINFHETENAAK
                     AMRKLLSGEINSIKLDNGDELKIKLNDEKHKDSTKWDKSYSFISNLEEEKYSQTDLFR
                     KKQEINEANTKIIEDRQEFYILNNDEIENIATRFVLENNFDELYIQSFKQSLIDIIQS
                     LNN"
misc_feature         8020. .10389
                     /note="possible cen1, region of very high [A+T] content"
gene                 14884. .20352
                     /gene="MAL1P3.03"
CDS                  14884. .20352
                     /gene="MAL1P3.03"
                     /note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
                     /codon_start=1
                     /product="putative ABC transporter"
                     /protein_id="CAB63558.1"
                     /db_xref="GI:6594246"
                     /translation="MTTYKENVGISNKGNKKKKSCQNISFLNFLSFDWIRPLINDLIK
                     GDIQELPNICRNFDVPYYASKLEENLRDIEVEDSEFYSEKNSSNEHVLHHCNSNDASE
                     KKVYNVYYHNILWSILKTFKFRIILIISFYILETLIVTLGGKFIDYYMRILEGQKIPV
                     YISFLKDFKVFSGLVVVMIMFFHLFPEALLHFYFHLFTINLKVSLMYFLYKINLCSNN
                     NHLQNPDAFYNTYRKFSSQTEIDEISRDFLSIGKNASSSSSGIKNNNKNIDNNKFVEN
                     DYIINFIKSTKKMEKDSLNENRSLPNVNIYNIMFSDVPSVTFFVTSCINLFNVFVKIF
                     MSFYVFHIKIGSNSVGIAIWLSIALYSAMILFEFLPSLFKSKYLIYRDKRIDNMHHVL
                     KEFKLIKMFNWESFAFKYINIFRMKEMKYCKIRLYLSNIGVFISSISSDIVEVVIFFI
                     YLKDRLNKKEEIKFTSIIMPLYVYKILISNVANFPNLVNNVMEGIVNIKRLNNYINDH
                     LYYNDIKNYFMYRTRYNEDYNIVVDKTFLQNENITSHDDGTSHNLKHLKNVIKNKLTN
                     MFKYFFFYHKMNYHKNIINKQILSGLLKNVDDNTNKKICFQEHKSNSTYNYNSSHIHE
                     KKEEYENIHNSSNSTMSNEFKEKKKNNEYIIKLENCSFGLSYDNKCDNDHILKNINFN
                     LKRNSLAIIIGNVGSGKSAFFHSILGDFNMTHGNLYIENFFKKMPILYVPQNSWLFMG
                     NIRSMILFGNEYNPLIYKYTILQSELLNDLSTIEHGDMKYINDDHNLSKGQKVRICLA
                     RALYEHYIHMHKLCTDYEKKLIQPNEILDKDLINNKNISSYNNKKSKLVNYNIPFNEN
                     YLQKCLMDDNNFYLYLLDDIFTSLDPSISKKIFSNLFCKEDNISFKDNCSFIISMNKS
                     TLDNFLIEDILDNVQYEVNIFEIQDKTLKYRGNISEYMEKNNLNITKESHWGYSNLNT
                     IDYTRIKLFDEVELNHVKHSNKMIYKEAYFVKGNTESVSFEIDSINKEYIKKMKKKNY
                     KKEHMNKNNKDNNNNNNSNKDDHININMNDNHRNYNDINLGPNSTDDSPTVSSLGNE
                     YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMPFVKSSSNMINNPSNFKYEDNSSS
                     FKGSISLETYLWYFQQVGFVLLTSVVIFMLISIFTDEIKFVFLTMMSIISKNNKEHSD
                     TILQKQVRYLEYFVILPIISLVTSGICFSMIIYGNITSAIKVHNNILYSILNAPLYIF
                     YNNNLGNIINRFIIDISAFDYGFLKRIYKAFFIFFRCILSSLLIIYMIRDCIFIFPFV
                     IILIYFFVFKRFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIYKKNTYHLDVYE
                     HYINNFRISYFFKWLINIWASLYIKIFILLLTTYIIMHPHLYASGIIKLYKEKNYVRI
```

```
                LSTLGYCISFSARLGVIIKFLLCDYTHIEKEMCCVQRLEEFAKISNKENASMNKENEL
                NVITTQTYKEKNENISDKISAIVEYKNVSLSSIINSSQDDESKKKYGIKFENVYVSYK
                KKIPLVNGTYKYIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQG
                KITVEGRDIRTYNRKGEDSIIGILAQSSFVFYNWNIRTFIDPYNNFTDDEIVHALKLN
                GINLGKNDLYKYMHKQDMKSNYKKIIQTSKVINQSNDNTILLTNDCIRYLSLVRLYLN
                RHKYKIILIDEIPIFNLNNSVHDELNSFLIGKAKSFNYIIRNHFPNNTVLIISHHANT
                LSCCDYIYVLRKGEITYRCSYEDVKTQSELSHLLEMDD"
rRNA            23896. .31533
                /gene="rRNA"
                /note="region containing small subunit, 5.8S and large
                subunit rRNA genes and spacer regions"
gene            23896. .31533
                /gene="rRNA"
gene            complement(join(31966. .32476,32675. .32775))
                /gene="MAL1P3.04"
CDS             complement(join(31966. .32476,32675. .32775))
                /gene="MAL1P3.04"
                /note="MAL1P3.04, conserved hypothetical membrane protein,
                len: 203 aa, similarity: P. falciparum chromosome 2,
                PFB0110W, O96126 predicted integral membrane protein (255
                aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
                in 191 aa overlap)"
                /codon_start=1
                /product="conserved hypothetical membrane protein,
                MAL1P3.04"
                /protein_id="CAB63559.1"
                /db_xref="GI:6594247"
                /translation="MKKSYTFINVTILLFLTLLLFLTYYNYDTFSKTKFNNNIKIDIN
                RFKRIIAEASEEQKYPWEEDFCLILNEEELIRPEHNDSPYLPEHYENIDKINELSINS
                TKIWKETIKKMRQNYEKETDNMNHNWRDFMWHYKWANIYLYKVHKLINITLKDLTNPI
                HDKEETITTWIKWIQEDIEYFLFNLQVEWLRILTLELFYKNKE"
misc_feature    complement(32477. .32486)
                /gene="MAL1P3.04"
                /note="potential splice acceptor sequence"
misc_feature    complement(32669. .32674)
                /gene="MAL1P3.04"
                /note="potential splice donor sequence, aaa/gtatat"
gene            join(36657. .36743,36864. .37343)
                /gene="MAL1P3.05"
CDS             join(36657. .36743,36864. .37343)
                /gene="MAL1P3.05"
                /note="MAL1P3.05, hypothetical protein, len: 188 aa"
                /codon_start=1
                /product="hypothetical protein, MAL1P3.05"
                /protein_id="CAB63560.1"
                /db_xref="GI:6594248"
                /translation="MRIKMNSGIFFIKLLICISFICVFECFNKCMISYRKDLLWYSEN
                CFNYSIDRSLAEGSSESKETKVKDIPNIELLKSLNINYEEYEKMKEIVGSFMDNNNLN
                IANEVLKNIHSFTNIENIFSLINDSSKSPVLKTFLKEFGSIFPHMLNNVPKLLFDLCQ
                RNPLHIILGLIVILAAIYVFENFKNFEC"
misc_feature    36744. .36749
                /gene="MAL1P3.05"
                /note="potential splice donor sequence, aag/gtatga"
misc_feature    36854. .36863
                /gene="MAL1P3.05"
                /note="potential splice acceptor sequence"
gene            complement(join(38049. .39995,40210. .40284))
                /gene="garp"
CDS             complement(join(38049. .39995,40210. .40284))
                /gene="garp"
                /note="MAL1P3.06, garp, len: 673 aa, similarity: almost
                identical to GARP_PLAFF (678 aa), fasta scores: 97.6%
                identity in 678 aa overlap"
                /codon_start=1
                /product="hypothetical garp protein"
                /protein_id="CAB63561.1"
                /db_xref="GI:6594249"
                /translation="MNVLFLSYNICILFFVVCTLNFSTKCFSNGLLKNQNILNKSFDS
                ITGRLLNETELEKNKDDNSKSETLLKEEKDEKDDVPTTSNDNLKNAHNNNEISSSTDP
                TNIINVNDKDNENSVDKKKDKKEKKHKKDKKEKKEKKDKKEKKDKKEKKHKKEKKHKK
                DKKKEENSEVMSLYKTGQHKPKNATEHGEENLYEEMVSEINNNAQGGLLLSSPYQYRE
                QGGCGIISSVHETSNDTKDNDKENISEDKKEDHQQEEMLKTLDKKERKQKEKEMKEQE
                KIEKKKKKQEEKEKKKQEKERKKQEKKERKQKEKEMKKQKKIEKERKKKEEKEKKKKK
                HDKENEETMQQPDQTSEETNNEIMVPLPSPLTDVTTPEEHKEGEHKEEEHKEGEHKEG
                EHKEEEHKEEEHKKEEHKSKEHKSKGKKDKGKKDKGKHKKAKKEKVKKHVVKNVIEDE
                DKDGVEIINLEDKEACEEQHITVESRPLSQPQCKLIDEPEQLTLMDKSKVEEKNLSIQ
                EQLIGTIGRVNVVPRRDNHKKKMAKIEEAELQKQKHVDKEEDKKEESKEVEEESKEVQ
                EDEEEVEEDEEEEEEEEEEEEEEEEEEEEEEEEDEVEEDEDDAEEDEDDAEEDEDDAEED
                DDDAEEDDDDAEEDDDEDEDEDEEEEEDEEEEEESEKKIKRNLRKNAKI"
misc_feature    complement(39996. .40005)
                /gene="garp"
                /note="potential splice acceptor sequence"
misc_feature    complement(40204. .40209)
                /gene="garp"
                /note="potential splice donor sequence, aag/gtaaca"
gene            join(45401. .46396,46562. .50233)
                /gene="MAL1P3.07"
```

```
Query Match              6.8%;  Score 80.2;  DB 3;  Length 67970;
Best Local Similarity   47.7%;  Pred. No. 0.00022;
Matches  335;  Conservative    0;  Mismatches  358;  Indels   10;  Gaps    3;

Qy      50 atgaacacaaagcaaagagaagttaggtcactcctcctttatatatatatatatatgcat 109
           || || | | |  |    |  | |    |  ||  |  |||||||||||||||||
Db    8027 ATTAATAAATATAATTAAAGAATATTTTATCTTTGTATTATATATATATATATTAAATG 8086

Qy     110 gcatgaggaccatggctatgatgaaggttaatagaggtagttgtgattgagatatgtcca 169
           | || |     ||   ||  || | |  | || | |     ||   ||  | | ||  ||
Db    8087 GAATAAATTAAATAAATAAAATAA---TAAAAATAATATATTAAAATATATAAATTAACA 8143

Qy     170 gcactagttttttgttggtgtgatttctcatgatgacgcgaaaaattttatatatatatat 229
             |  |   | |  ||       ||  ||       ||||   ||| |   | ||
Db    8144 AAATAATAATATAATTAAATAAATAAAATATTATATTAAATAAATAAAATAAAATAAAAT 8203

Qy     230 aatgaataatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaat 289
           | | || |||||  | | | || | |   |||||   | |||||    || | |
Db    8204 ATTAAAAAATATAAATTAATAATATAATTAATTAAATAAAATATTAAAATATATAAATTA 8263

Qy     290 gtgaggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcac 349
           | |  | |  | |    |||  |  |    |  |      |    | | || ||||
Db    8264 ATAATATAATTAATAAATAAAATATTATATTAAATAAATTAAATAAAATATTAAAATATA 8323

Qy     350 aactcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggtta---- 405
             |    | |   | | | ||   | |||| |||     ||||||| | |    ||
Db    8324 TAAATTAATAATATAATTAATAAATAAAATATTATATATATTAAATAAATTAAATAAAAT 8383

Qy     406 attcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaa 465
           |||  |    |          | ||| |||||| || |  ||| ||| ||  |||| ||||
Db    8384 ATTAAAATATATAAATTAATAATATAATTAATAAATAAAATATTATATTAAATAAATTAA 8443

Qy     466 ttaaaattttattggatttgaatgaactcaat---tacatcacaaaaaacctaatcaaat 522
           |||||| |||    || | ||| ||    ||   ||  | |  ||| |    || | ||
Db    8444 ATAAAATATTAAAATATATAAATTAATAATATATATAATTAAATAAATAAACTATTATAT 8503

Qy     523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
           |||      |   | |||||||| | ||||| ||     ||     |||||
Db    8504 TAAATAAATTAATAATATAATTAATAAATAAAATATTATATTAAATAAATTAAATAAAAT 8563

Qy     583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggccagg 642
             | | | |  |  ||  | |       ||| || |  ||  ||   |  |
Db    8564 ATTAAAATATATAAATTAATAATATATATAATTAAATAAATAAAATATTATATTAAATAA 8623

Qy     643 acaattattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaa 702
           |  | || || |||||||||| ||  || || | | | | |         || | ||||||
Db    8624 ATTAATAATATAATTAATAAATAAAATATTATATTAAATAAATTAAATAAAATATTAAAA 8683

Qy     703 tatatgtctaaacactaataataaaatttatttgtgtatcttt 745
           |||||   | || | ||  | ||| |  ||      ||| | |
Db    8684 TATATAAATTAATAATATAATTAATAAATAAAATATTATATAT 8726


RESULT   9
AL513330/c
LOCUS       AL513330              202645 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-469E8, *** SEQUENCING IN
            PROGRESS ***, 7 unordered pieces.
```

```
ACCESSION   AL513330
VERSION     AL513330.9  GI:13990571
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202645)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13396737.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA469E8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 0% of reads Chemistry:
            Dye-terminator Big Dye; 99% of reads
            Consensus quality: 200655 bases at least Q40
            Consensus quality: 201365 bases at least Q30
            Consensus quality: 201775 bases at least Q20
            Insert size: 202045; sum-of-contigs
            Insert size: 161399; 19.6% error; agarose-fp
            Quality coverage: 6.15x in Q20 bases; sum-of-contigs Quality
            coverage: 7.70x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7983: contig of 7983 bp in length
            *     7984 8083: gap of      100 bp
            *     8084    16795: contig of 8712 bp in length
            *    16796 16895: gap of      100 bp
            *    16896    73805: contig of 56910 bp in length
            *    73806 73905: gap of      100 bp
            *    73906   147587: contig of 73682 bp in length
            *   147588 147687: gap of      100 bp
            *   147688   178901: contig of 31214 bp in length
            *   178902 179001: gap of      100 bp
            *   179002   181519: contig of 2518 bp in length
            *   181520 181619: gap of      100 bp
            *   181620   202645: contig of 21026 bp in length.
FEATURES             Location/Qualifiers
     source          1. .202645
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-469E8"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .7983
                     /note="assembly_fragment:00874
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8084. .16795
                     /note="assembly_fragment:01844
                     fragment_chain:1"
     misc_feature    16896. .73805
                     /note="assembly_fragment:01698
                     fragment_chain:1"
     misc_feature    73906. .147587
                     /note="assembly_fragment:03055
                     fragment_chain:1"
     misc_feature    147688. .178901
                     /note="assembly_fragment:01073
                     fragment_chain:1"
     misc_feature    179002. .181519
                     /note="assembly_fragment:01557
                     fragment_chain:1"
     misc_feature    181620. .202645
                     /note="assembly_fragment:01958
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
BASE COUNT    67892 a  34283 c  34403 g  65461 t    606 others
ORIGIN


  Query Match             6.8%;  Score 80.2;  DB 2;  Length 202645;
  Best Local Similarity   49.9%;  Pred. No. 0.00017;
  Matches  235;  Conservative    0;  Mismatches  228;  Indels    8;  Gaps    1;

Qy     336 atttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaat 395
           || ||| | ||    | | || || |||  || |  | ||| | || |  | |||  ||
Db 202524 ATATTTAATATATATATTTAAATTTAAATATATATATTAAATATATATTTAATATATAAA 202465

Qy     396 tattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtattttt 455
           |||    || |   |    ||   ||   || |||   |  | || |      ||| | |
Db 202464 TATATATTTAAATATATATTTAATATATAAATATATATTTTAAATATATATTTATATAT 202405

Qy     456 gataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaaaaccta 515
           |  || || ||||| || |||    ||| ||  |  |   ||| |  | | |     ||
Db 202404 AAATATATATTTAAATTTATATACATTTTTAAATATATATTTTATACAAAATATTTTATA 202345

Qy     516 atcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta------ 569
           ||||| |||| || |   |    || ||   | |||  |    ||  | ||
Db 202344 TAAAAATTTATATATTTTTAAAAGATATATATTTAAATATTTTTAAAATATATATATTTT 202285

Qy     570 --aatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctatta 627
             ||| |  | ||| | ||||||     || ||  ||         || ||| |  ||| |
Db 202284 ATAATATATAATTTATATTATAATGTGTACATAATATATATTATAATATAATATATATAA 202225

Qy     628 ttcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcact 687
           | | |  |             ||   | | | ||   ||||| || | || || || |    |
Db 202224 TACTATATATTATATTATATATTATAATATATATTATAATATATTATATTATACATAT 202165

Qy     688 tcctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttgg 747
           |   |  | ||| |||||||| |   |  | || |||| || ||||    ||| | |
Db 202164 TATAATGTATTATAATATATATTATATTATATATTATAATATATTATATTATATATTA 202105

Qy     748 cagtaggtgagaggtgctgacaaataaattagtgcataaaatataatggat 798
           | ||  |   |  | |   | ||| ||||    | | | ||||||  ||
Db 202104 TAATATATATTATATTATATTATATATTATAATACATATTATAATACAT 202054


RESULT  10
AC096550
LOCUS       AC096550              181486 bp    DNA     linear   HTG 18-SEP-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-20F13, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC096550 AC013666
VERSION     AC096550.1  GI:15638715
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181486)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181486)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:7382143.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: H_NH0020F13
            Drafting center: WIBR
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 46%
            Sequencing vector: plasmid; 54%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 176334 bases at least Q40
            Consensus quality: 178430 bases at least Q30
            Consensus quality: 179303 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1290: contig of 1290 bp in length
            *     1291     1390: gap of unknown length
            *     1391     2839: contig of 1449 bp in length
            *     2840     2939: gap of unknown length
            *     2940     4962: contig of 2023 bp in length
            *     4963     5062: gap of unknown length
            *     5063     6591: contig of 1529 bp in length
            *     6592     6691: gap of unknown length
            *     6692     8921: contig of 2230 bp in length
            *     8922     9021: gap of unknown length
            *     9022    11487: contig of 2466 bp in length
            *    11488    11587: gap of unknown length
            *    11588   181486: contig of 169899 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-20F13"
     misc_feature    1. .1290
                     /note="assembly_name:Contig2"
     misc_feature    1391. .2839
                     /note="assembly_name:Contig3"
     misc_feature    2940. .4962
                     /note="assembly_name:Contig4"
     misc_feature    5063. .6591
                     /note="assembly_name:Contig5"
     misc_feature    6692. .8921
                     /note="assembly_name:Contig6"
     misc_feature    9022. .11487
                     /note="assembly_name:Contig7"
     misc_feature    11588. .181486
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    59725 a  33010 c  32809 g  55334 t    608 others
ORIGIN


  Query Match              6.7%;  Score 78.8;  DB 2;  Length 181486;
  Best Local Similarity   45.9%;  Pred. No. 0.00029;
  Matches  269;  Conservative    0;  Mismatches  317;  Indels    0;  Gaps    0;

Qy     213 attttatatatatatataatgaataatatgattgattattctctgtaattttgtgaaata 272
           || ||||||| |||||| |  || |||| |   |||||      ||||| | | | ||
Db  103963 ATATTATATAATTATATATTATATTATATAAAAAATTATATATAATAATTATATTATATT 104022

Qy     273 gattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggcatg 332
           |||| |     | |||   |    |  |  |    |   | || |  |  |     ||
Db  104023 AATTACATATAATTAATTATATAATATAATATAATTATATAACATTATATATTATATATA 104082

Qy     333 gtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattatta 392
           |   |  | | || | |  | |  ||  |   ||  ||  | |     ||| ||
Db  104083 ATATATAATTATATTATACATTATATTATTATTATTATATATTATTATATATAAT 104142

Qy     393 aattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtatt 452
           |  ||||| |||  |      |    ||   | |  |||||  || |       | |
Db  104143 ATATATTATATTATGTATATTATGTATTATATATTATATATTATGTATATTATATATTAT 104202

Qy     453 tttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaaaac 512
           ||    | |||  || ||||   ||  | || | |  |   |||| || | | | ||
Db  104203 ATTATATAATTATATATAATTAATTTATAATTAATTATATATAATTATATTATATAATAAT 104262

Qy     513 ctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaat 572
           ||||  |  |  ||| |     |     |||   |||   |    ||   | ||| |
Db  104263 ATAATGTATATGTTATATATAATATACAAATAATATAATTATATATTATATAATATAATT 104322

Qy     573 ctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaa 632
           |  | | |  |||  |||  |  ||     |    || |||   ||  |||||  ||
Db  104323 ATATATTATATAATATGTAAATAATATAATTATATATTACATTATTCACATATTATATAA 104382

Qy     633 actggccaggacaattattaaaattaataattattatttttttctaataaagcacttccta 692
           |          ||| |  |   | |||  || ||| | || ||  | |     |   |
Db  104383 TATATAATTATATATTCTATATTATTATATATAATATATATTATATAATATATAGTATAA 104442

Qy     693 attgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagta 752
           | | |  ||||||     |  | ||  ||||  || ||||     || | |  |  |
Db  104443 TATATAATTATATATTCTATATTATTATATATAATATATATTATATAATATATAATAATAA 104502

Qy     753 ggtgagaggtgctgacaaataaattagtgcataaaatataatggat 798
           |   |     |     |||   |  |  |||| |||||  |  ||
Db  104503 TATATAATTATATATTCTATATTATTATATATAATATATATTATAT 104548


RESULT  11
AE001398/c
LOCUS       AE001398               14867 bp    DNA     linear   INV 06-NOV-1998
DEFINITION  Plasmodium falciparum chromosome 2, section 35 of 73 of the
            complete sequence.
ACCESSION   AE001398 AE001362
VERSION     AE001398.1  GI:3845197
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 14867)
  AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
            Smith,H.O., Fraser,C.M., Hoffman,S.L. et al.
  TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
  JOURNAL   Science 282 (5391), 1126-1132 (1998)
  MEDLINE   99021743
  REMARK    Erratum:[[published erratum appears in Science 1998 Dec
            4;282(5395):1827]]
REFERENCE   2  (bases 1 to 14867)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
            Medical Center Drive, Rockville, MD 20814, USA
FEATURES             Location/Qualifiers
```

```
     source          1. .14867
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="2"
     gene            complement(1570. .2424)
                     /gene="PFB0490c"
     CDS             complement(1570. .2424)
                     /gene="PFB0490c"
                     /note="predicted by GlimmerM"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71887.1"
                     /db_xref="GI:3845198"
                     /translation="MKEKNEKIMDYLSCPLDDVVDREKKSGKNSLLKSSSTKKSDYKK
                     SSIFSKKRDSHKKGSSFRGRRSGFINRKSGSFKKPYYNNRLINKNYNNYKGRNFHNGR
                     DNFKGRTGSFGSRVFDNRKGSFKKRFISNRNKSSVKSYRGNGSNKMGRKSFNKAPTSR
                     TVVTKRLNNYKTVSAPVKKFNNLNISLYRKNRTFALNTKRSKPVGTIKSSVPRKRIKK
                     GLKKGSLKSKTRKSTSGSKFKPLNKYFLSKIKIVTSLNKIPSPLKEQKNTEVNLPESL
                     NNATTKKN"
     gene            11241. .14606
                     /gene="PFB0495w"
     CDS             11241. .14606
                     /gene="PFB0495w"
                     /note="predicted by GlimmerM"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71888.1"
                     /db_xref="GI:3845199"
                     /translation="MFALKKNTVREGFVNICFSYLKKLYLKSNFVTVNLNYETNNEKR
                     TNKKIYKKSKAQSLFDKGLNIHDKLILFKNLPKYKCAKYECISAKEVYKYLLDEYKKC
                     FNYISLCDIIQSVKIFDELDKTFTDYNFYIEVKNIDKNVLNKINEIYFKNKDITFHRR
                     EILGKICNKIMSYIHEMNGNELIHFLIYFFRWNKNDKNLILFYNYYFNYVFDHMYLFN
                     HEIYKLLFIFNKYLNNNSNIPFNKNLIQEMEFNLYYFREIKNEKNYIIKMNKKEIYKK
                     CFAKFHENVDHIDNEKILNILRLYVDNSILDIDINNKMLCNLNNNLINENIEYISKLL
                     NFYCTLIKKGKYDNDMTIYKLKEVIKATHHILCDKTKNLETFCSDIDYSTLLNSLNNK
                     FILNKIIDKNFILFYECLLKILLNIKFVNFQSLCISLISLKNIYYNILRNNVYIVNNV
                     LFNDIMKFSLYLCNIFLGKRIKTENENAVLIIHNNDQTNYSNKENIKDIIIQKRIKEY
                     IFYKMENYKDFHFKLKDSDLLSIKLLSNTFVKINEVYNSYDFYLLFNNISCILYNFLV
                     NRNSVKKYKDTYIYILNDLSFVYKYIKNNDRTKKKKNFFLLSSSMKELICKNILSVSN
                     RYIKHLHEEDNFDQKDQYVCSLTFLNNLFFDKIIHFHYIYNLWCHVYKTYNYFKCNKL
                     INEDIISLLLLTCSKFQYFIENNSNDRYCRKELIHLKYNIIDDLIKNYLNTYKSISID
                     NISKIFISLSNSKYTCEVNENLLLESLQSEFEKVTKTSKKGGIHMMDNNLLDNNNSCE
                     KYEHRYIEYKKENLFINLNKIIECLIKLNIFLYLKKKKTYLYLYKQSLCPINLKENIL
                     KKILYIANNLYMYEMYGYVCEMLERVLSSHKEQNLFSYNYNKNVEHKMFDKILCHISE
                     DDYIEMSNTMYVLFYDYLKNINSERQSNILRNNSTNDRFIDEIKEKKYKLNNNTLIKH
                     NNVKLNYEKSNNSNGNISNILKDDKNKNHNNVEMDLIDNKNENKKIQEKGQNGENCEN
                     CKDVLVNDIINIFGFLKMEKKKFLFFQLYMYLCNITKFKRRYVSSSSLFHMDVFKIIK
                     DMNLKYLCLENYKIKNEECAFLYTIDIVLFKER"
BASE COUNT     6284 a   1019 c   1106 g   6458 t
ORIGIN


  Query Match             6.7%;  Score 78.4;  DB 3;  Length 14867;
  Best Local Similarity   48.1%;  Pred. No. 0.00061;
  Matches  260;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    1;

Qy    210 aaaatttttatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
          || |  | | |||    ||||||   || ||      || |||   | ||||  | | ||
Db   8497 AATAAATGAAATAATATTATAATTTTTATTAATTAATATAATTTAATTTAATAATATAAA 8438

Qy    270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
          | || | || |  | |||   | |  || |  | |   || |     |  | ||
Db   8437 TAAAATAATAATATATTAATACAAAGAAACAAAATATATTATTAAATATATAATTTTAAT 8378

Qy    330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
            | ||| |  |||  | ||   |       |   | |||  | ||  |||||  | |
Db   8377 TAATTAATTATAAAAAATAAAATAAATAGACTCATAAATAATATATAATAATTTAATTAA 8318

Qy    390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
            ||    |||   || ||      ||   |||| |||||    ||  | ||       |
Db   8317 AAAATAATTTAATATATTTTTAATTTTAATTAATTAATTAAATATATATTAATATTATAT 8258

Qy    450 atttttgataatttaattaaaattttattggatt---------tgaatgaactcaattac 500
          |||| | |||||||||||  |  |||| |   | |         | ||| || |  ||||
Db   8257 ATATATCATAATTTAATATATTTTTTTTATAAATAATAATTAATTAATTAAGTTTATTAA 8198

Qy    501 atcacaaaaaacctaatcaaattaatatcttatgtgatataatttagaaatataaatgat 560
          |    || | |   | | || ||||     |||     ||| || | | | || ||
Db   8197 AAATAAATATATTAATTTAATTTAAATATATATAAATAATATTTAATATAAATTAAATTA 8138

Qy    561 taacctttaaatctcgagtttctcttataaaaaacacgtataattgggctagatttaaca 620
            |  |||| | |  | | |   ||  || ||| |   |||| |    || || ||| |
Db   8137 ATATAATTAATTATTTATTATTCATTTAAATAAAAAATAATAAATTAAATATATATAATA 8078

Qy    621 gctattattcaaactggccaggacaattattaaaattaataattattatttttctaata 680
            |||  || | |          |  || |||| ||||| || || |  ||  ||     |
Db   8077 TATATATTTTATATGTAAATATATTATAATTATAATTATTATTTTTATTTAATTAATTAA 8018

Qy    681 aagcacttcctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgta 740
              ||  || ||  | |  ||| |  |     |  ||||| || | ||| | | ||
Db   8017 TTTATATTATTATTTTATTATTTATTTAACATTTAAAAAATAAATAATTATATATATATA 7958


RESULT  12
AC074035
LOCUS       AC074035               189214 bp    DNA     linear   HTG 30-NOV-2001
DEFINITION  Homo sapiens chromosome Xp clone RP11-707P20, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC074035
VERSION     AC074035.21  GI:17149341
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189214)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189214)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 29, 2001 this sequence version replaced gi:16327947.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBYI
                Center clone name: RP11-707P20
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 190931 bases at least Q40
                Consensus quality: 191911 bases at least Q30
                Consensus quality: 192818 bases at least Q20
                Estimated insert size: 190344; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 17.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   130843: contig of 130843 bp in length
            *  130844   130943: gap of unknown length
            *  130944   189214: contig of 58271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189214
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xp"
                     /clone="RP11-707P20"
BASE COUNT    57784 a  35775 c  35259 g  60296 t    100 others
ORIGIN


  Query Match              6.7%;  Score 78.4;  DB 2;  Length 189214;
  Best Local Similarity   46.6%;  Pred. No. 0.00034;
  Matches  250;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy    210 aaaattttatatatatatatatgaataatatgattgattattctctgtaattttgtgaa 269
          | ||  |||||||||| |||||   ||||| || ||  || | | | | ||||    | |
Db  53852 ATAAAATTATATATAAATATATATAATAAAATTATATATAAATATATATAATAAAATTAT 53911

Qy    270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
          |||  |   | | |    ||  | |  |    |  | | | | |  |
Db  53912 ATATAAATATATATAATAAAATTATATATAAATATATAATAAAATTATATATAAATATAT 53971

Qy    330 atggtgattttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
          ||  | |  ||  | || |  |   |||   | ||    |  || |||    | | || |
Db  53972 ATAATAAAATTATATATAAATATATAATAAAATAAAATTACATATAAATATATAATATAA 54031

Qy    390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
          | |||||||       ||       ||   ||   |||       ||||| |   | | |
Db  54032 TAAAATTATATATAATATAATAAAATTATATATATAATATAATAAAATTATATATAAAAT 54091

Qy    450 atttttgataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaa 509
          |  |  |   |  || || || | ||   ||   | |  |    | || || | | ||

Db  54092 TTAATAAAATTATATATCAATATATAATATAATAAAATTATATATCAATATATAATATAA 54151

Qy    510 aacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta 569
          |       | || ||||| | ||   ||   || |  | ||||||   | |||  ||
Db  54152 TAAAATTATATATCAATATATAATAAAATTATATATCAATATATAATATAATAAAATTAT 54211

Qy    570 aatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattatt 629
          |     | |    ||  ||||  |  ||| | |    || || |   |  ||||  |
Db  54212 ATATCAATATATAATATAATAAAATTATATATCAATATATTATATATGTAATATATTTAT 54271

Qy    630 caaactggccaggacaattattaaaattaataattattatttttttctaataaagcacttc 689
            | |    | |    ||||| |  || | ||  |  ||| |    |||||    ||
Db  54272 ATTATTATATATAATTTATATTACATATATTTATATTATTTTATATGTATAAATATATTA 54331

Qy    690 ctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatcttt 745
            |||   | | |||| | | |||| |  |   ||  || | | |     || |||
Db  54332 TAAATATGTGATATATTTATGTAAATATGATATATTGAAATAAATATACCATATTT 54387


RESULT  13
AC108683
LOCUS       AC108683               208319 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome Xp clone RP11-539J11, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC108683
VERSION     AC108683.1  GI:18449744
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208319)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208319)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDOL
                Center clone name: RP11-539J11
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 211007 bases at least Q40
                Consensus quality: 211494 bases at least Q30
                Consensus quality: 211917 bases at least Q20
                Estimated insert size: 210102; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   171542: contig of 171542 bp in length
            *  171543   171642: gap of unknown length
            *  171643   196472: contig of 24830 bp in length
            *  196473   196572: gap of unknown length
            *  196573   208319: contig of 11747 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xp"
                     /clone="RP11-539J11"
BASE COUNT    64557 a  40319 c  37431 g  65812 t    200 others
ORIGIN


  Query Match              6.7%;  Score 78.4;  DB 2;  Length 208319;
  Best Local Similarity   46.6%;  Pred. No. 0.00033;
  Matches  250;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy     210 aaaatttatatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
           | ||  |||||||||| |||||   ||||| || ||  || | | | | ||||    | |
Db  188568 ATAAAATTATATATAAATATATATAATAAAATTATATATAAATATATATAATAAAATTAT 188627

Qy     270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
           ||| |   | | |    ||  | |   |    |  | | | |  |  |
Db  188628 ATATAAATATATATAATAAAATTATATATAAATATATAATAAAATTATATATAAATATAT 188687

Qy     330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
           ||  | |  ||  | || |  |   |||   | ||    |  || |||    | | || |
Db  188688 ATAATAAAATTATATATAAATATATAATAAAATAAAATTACATATAAATATATAATATAA 188747

Qy     390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
           | |||||||       ||       ||   ||   |||       |||||  |   | | |
Db  188748 TAAAATTATATATAATATAATAAAATTATATATATAATATAATAAAATTATATATAAAAT 188807

Qy     450 atttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaa 509
           |  |  |    |  || || || | ||   ||   | |  |    | || || | | ||
Db  188808 TTAATAAAATTATATATCAATATATAATATAATAAAATTATATATCAATATATAATATAA 188867

Qy     510 aacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta 569
           |       | || ||||| | ||   ||   || |  | ||||||   | |||  ||
Db  188868 TAAAATTATATATCAATATATAATAAAATTATATATCAATATATAATATAATAAAATTAT 188927

Qy     570 aatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattatt 629
           |       | |    ||  ||||  |  ||| | |    || || |   |  ||||  |
Db  188928 ATATCAATATATAATATAATAAAATTATATATCAATATATTATATATGTAATATATTTAT 188987

Qy     630 caaactggccaggacaattattaaaattaataattattatttttttctaataaagcacttc 689
           | |    |  |    ||||| |  || | ||  |  ||| |    |||||    ||
Db  188988 ATTATTATATATAATTTATATTACATATATTTATATTATTTTATATGTATAAATATATTA 189047

Qy     690 ctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatcttt 745
           |||   | | |||| | | |||| |  |   ||  || | | |    || |||
Db  189048 TAAATATGTGATATATTTATGTAAATATGATATATTGAAATAAATATACCATATTT 189103


RESULT  14
PFMAL3P5
LOCUS       PFMAL3P5               86827 bp    DNA     linear   INV 04-MAY-2000
DEFINITION  Plasmodium falciparum MAL3P5, complete sequence.
ACCESSION   AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
            AL010206 AL010210 AL139179
VERSION     AL034556.3  GI:7711064
KEYWORDS    HTG; centromere; CTRP protein; initiation factor E4;
            Serine/threonine protein phosphatase.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 86827)
  AUTHORS   Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
            Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,
            Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
            Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
            Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
            Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
            Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
            Barrell,B.G.
  TITLE     The complete nucleotide sequence of chromosome 3 of Plasmodium
            falciparum
  JOURNAL   Nature 400 (6744), 532-538 (1999)
  MEDLINE   99376085
REFERENCE   2  (bases 1 to 86827)
  AUTHORS   Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quail,M. and
            Barrell,B.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 86827)
  AUTHORS   Lawson,D., Bowman,S. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On May 14, 2001 this sequence version replaced gi:2982535
            gi:2982536 gi:2894454 gi:2982554 gi:2982562 gi:2894489 gi:2982572
            gi:2982574 gi:4493931.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .86827
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="MAL3P5"
     gene            join(322. .603,826. .1023,1197. .1301,1458. .2942)
                     /gene="PFC0575w, MAL3P5.1"
     CDS             join(322. .603,826. .1023,1197. .1301,1458. .2942)
                     /gene="PFC0575w, MAL3P5.1"
                     /note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689
                     aa, possible signal sequence, revised: added new exon 2"
```

/codon_start=1
/product="Hypothetical protein, PFC0575w"
/protein_id="CAB38969.2"
/db_xref="GI:7711065"
/db_xref="SPTREMBL:O97258"
/translation="MYLKNVYIYISSCFILFDLCFSFHLLKMKYKNHMNNMKSVTFFL
RSPQIYRKRFKRSRIKNVSFKKKQKKPLFLFENLKKGFSFLGFWRNQYDQKYIDDVIS
NINNLTRIKQVTHKKKSNEFTKENIKQILLHCVFSKIDFKIINNLSYIIKHFQMSNIT
VHSILNQISEKVKEKKDAENYLALHLFLLKDENITLFSMMHIMDFFKSKQKVIECIRD
IKSKKKRKKNLSIYINLFICTLIYFTYCMCLLIKYISHLCIFFFFFFCFFLCYNILER
IYEECVGDLIRKKIERYNLYCEKKKIKFHMKDAIKKMEINMKDDDLYFNYHYDELLRC
FTMKLNIERNNKNIIRSNYDNINNDISIDKDMYMNNPIDVNINNISLDEKIKEQFENP
DDENLKELKDTYEQFQLFNDNIIKYIEEDQPLYNINDNSNINDNNNNINTMKNKHKIK
DTYNDDDDYDYEKEEDLVIQKNIDDYIYKNTIGMNKSLEEFKNQFIEQADIEFQNFLS
NVNLDQHGRVKSNDENTKSTEHIKNKNTINKGYDTELIQNQMENNFIKKNIDNNISND
ISNNEINIIKLKKLNQSDEDINLTSDLIYERLRTKVLWYIQKIEYLKFKYQYDIINEQ
YPIIKNEKTVLDLLNYGYKIVMSPDVDNSLFEKTKIDSIPNEKDKNNQMENQKNSKNY
IIMFNVM"

misc_feature 604. .609
/gene="PFC0575w, MAL3P5.1"
/note="predicted splice donor sequence for exon 1 of PFC0575w (revised)"

misc_feature 816. .825
/gene="PFC0575w, MAL3P5.1"
/note="predicted splice acceptor sequence for exon 2 (revised of PFC0575w)"

misc_feature 1024. .1029
/gene="PFC0575w, MAL3P5.1"
/note="predicted splice donor sequence for exon 2 of PFC0575w (revised)"

gene complement(3354. .6644)
/gene="PFC0580c, MAL3P5.2"

CDS complement(3354. .6644)
/gene="PFC0580c, MAL3P5.2"
/note="PFC0580c (MAL3P5.2), Hypothetical protein, len: 1097 aa, possible signal sequence, predicted using hexExon"
/codon_start=1
/product="Hypothetical protein, PFC0580c"
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGLKRKNVFYLLVSVPSLFAYFLKRHKDNENNYETLINNNDIE
KIKKIRIHNKCSYIPLLFLNIYDSYIYKNKILRWLYFKFRKRRKDKEEYYYITNMVRK
KRREAIKYNFISDEQNLFNKFYIYEIVLEYSLKYGILSPHLSLYILKNISEHCVNIYP
SLYYYNKLDNKHNLINEKKLKYFKQINNEHTQQAPTNHTHHNNNNNNKKPLDINIHSC
KNTNISSYSTYNNMEKENINIYDKYNIHNFYTEKSISYKDENCQHITLNMIYLLNQTY
DNICRICLNTNTNIYINFYMINILKYICYKNMEIILLNYNHIEDMKKKINQKNNTNTS
LFKYIYSFFFFKKEENHIYDLFEDQMMNHLHKKENDKFYNYSNENTHNNIYKYISDNY
FYDHINSSSNRCSFKNLKKQQTDDNTKHIIMGKEKYPMNKSDHEKKNNNTCGNINIEK
DQKKDILKKIYFLKGNKLDDIQILNELYVMIYMRLLFECSLKLISIKKNIHLLEKKME
FDKDNKIIYLNSADYMNNLRRNILKRFSKNEERENINSFASFPFLLSKNIIYFEDEIG
RSRDNTIYNNVYDKETNKTTTNNNNNDNNDNICSNNDHICSNNNDHICSNNDHICSN
NNNNICSNNNNNICSNNNNNICSNNNNNICSNKMLDEFCQDNKFNDYNTRKKEKRKRI
YELAKIYTNNIFDYLKGKKEKHQNEDNTINLYYIKKKFPWIFYLKNIIKNKDTSFIEH
NNNIVNGDIKNNNIIFKKKYNLFESSIISYFYIKDIYEYNYKLRLYYIYDNLIKKFCR
YFLKMNEHINRKLYKMKRAFHYYIYNFDQFIINNYYHIIHKKNIHKIHIHLKQCKDKE
IDIVKFKDLYYCMINNINNIFSYIHKVDHNECVYRIFKAYNKILLYEYNYLNEKENIY
YKNKIKKYLTYLNNNISNDLYPYNISYNKIYNQNKYKNRKNFSHIFYSLKNDIHLLLF
LYTQRIQNCCDIFSYIYKKYNFNEKNPFLNYLYYELHYIVYSEKKKKKKFFSFISSSP
YSYDTMVNSFTFSYFFFSLSYLLFILFYHPDMYASYIFFKTLTYSGLPTYYYSLYNNI
MVVCGPKT"

gene join(7669. .7717,7800. .7829,7912. .7940,8064. .8106, 8302. .8368,8568. .8641,8713. .8812,8924. .9006,9122. .9258, 9369. .9505,9613. .9838)
/gene="PFC0581w"

CDS join(7669. .7717,7800. .7829,7912. .7940,8064. .8106, 8302. .8368,8568. .8641,8713. .8812,8924. .9006,9122. .9258, 9369. .9505,9613. .9838)
/gene="PFC0581w"
/note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction very tentative"
/codon_start=1
/product="hypothetical protein, PFC0581w"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/translation="MKKKKKHYFSIKFVNVDKNKYVLCSKDYIRIINYMIGLHIFRLL
QKIFHIYFANEQIASSFFESHKNYRVTKEDIIDGIEKCWFNITDYLISESIKQDNDFS
NDIKTTVTAMKNKMDQLLTTSYSNKKIDTVNASFQWAQSPEYIFLNIKFSHRWSSPGA
LKVKDEEIVSKKNNFSFSALSNDSNSVTKKYIVDLTLLDNIIESETKYNFASVGKVVV
TLKKEKKKIWNRLLLSKEKYPNMQVWWDMKEKRIHIITFVTINLFFLLSLSHRYHDSV
QNFLKEEKNNSDKLQDDIDEDEEKYFDEEILREAKKKSEEYDKDDEEL"

misc_feature 7790. .7799
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 2 of PFC0581w"

misc_feature 7830. .7835
/gene="PFC0581w"
/note="potential splice donor sequence for exon 2 of PFC0581w"

misc_feature 7901. .7911
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 3 of PFC0581w"

misc_feature 7941. .7947
/gene="PFC0581w"
/note="potential splice donor sequence for exon 3 of PFC0581w"

misc_feature 8055. .8063
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 4 of PFC0581w"

misc_feature 8107. .8112
/gene="PFC0581w"
/note="potential splice donor sequence for exon 4 of PFC0581w"

misc_feature 8293. .8301
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 5 of PFC0581w"

misc_feature 8369. .8374
/gene="PFC0581w"
/note="potential splice donor sequence for exon 5 of PFC0581w"

misc_feature 8559. .8567
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 6 of PFC0581w"

misc_feature 8642. .8647
/gene="PFC0581w"
/note="potential splice donor sequence for exon 6 of PFC0581w"

misc_feature 8704. .8712
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 7 of PFC0581w"

misc_feature 8813. .8818
/gene="PFC0581w"
/note="potential splice donor sequence for exon 7 of PFC0581w"

misc_feature 8913. .8923
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 8 of PFC0581w"

misc_feature 9007. .9012
/gene="PFC0581w"
/note="potential splice donor sequence for exon 8 of PFC0581w"

misc_feature 9112. .9121
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 9 of PFC0581w"

misc_feature 9259. .9264
/gene="PFC0581w"
/note="potential splice donor sequence for exon 9 of

```
                PFC0581w"
     misc_feature    9361. .9368
                     /gene="PFC0581w"

  Query Match             6.7%;  Score 78;  DB 3;  Length 86827;
  Best Local Similarity   49.3%;  Pred. No. 0.00047;
  Matches  265;  Conservative    0;  Mismatches  265;  Indels    8;  Gaps    2;

Qy     210 aaaatttatatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
           | || || ||||||||   | || || |  | ||  |||| | | | | ||||  | ||
Db   39326 ATAAATTAATATATATTATTTATTAAAATAAAAATAAATTAATATATATTATTTATTAAA 39385

Qy     270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
           ||| |   |||    |  || | |  |    |  |   ||   |  | | |  |
Db   39386 ATAAAAATAAATTAATATATATTATTTATTAAAATAAAAATAAATTAATATATAT----- 39440

Qy     330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
           |  | |||     ||  | || | ||| |  |   |  | | ||| || | | |  | |
Db   39441 -TATTTATTAAAATAAAAATAAATTAATATATATTATTTATTAAAATAAAAATAAATTAA 39499

Qy     390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
           | |    | ||  | |||| |    |  | |   | ||| | |||| || |    | |
Db   39500 TAATTAAAATAAATAAATTAATATATATTATTAATTAATATAAATAATAAATAAATAATT 39559

Qy     450 atttttgataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaa 509
           |   |  |||| |||||  | ||| |     |    ||  ||   |   |       |
Db   39560 AAAATAAATAAATTAATATATATTATTAATTAAAATAAATAATAAATAAATTAATAATTC 39619

Qy     510 aacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta 569
           ||   |||  | |  |||   ||| | |  ||| |||  ||| ||||| | |||   ||
Db   39620 AAATAAATATATTATATAATATATATTAATTAAATTAATAATTTAAATAAATAAATGTTT 39679

Qy     570 aatctcgagtttctcttataaaaaacacgt--ataattgggctagatttaacagctatta 627
           | |  | |   | ||||||  ||| |  |  ||||||    |  || ||| |  |||||
Db   39680 TTTATTAATTAAATTTTATAATAAATAAATTAATAATTCAAATTAATATAATATATATTA 39739

Qy     628 ttcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcact 687
           | |||      |    |    ||| ||| |||||  |||  |  |    ||  ||  | |
Db   39740 ATTAAATGAATTAAATAAGCAGTTATAATAAATAAATATATTAATAATAAAATAAATAAT 39799

Qy     688 tcctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatcttt 745
           |  | ||| |       |   | || | | ||| |  | |||| | ||   ||   ||
Db   39800 TTATTATTTTATTCCACTTAATATATATATTAAAATAATAATTAAATTAAATAATATT 39857


RESULT  15
AC026698/c
LOCUS       AC026698               192731 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Homo sapiens chromosome 5 clone CTC-457H17, complete sequence.
ACCESSION   AC026698
VERSION     AC026698.6  GI:18425273
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 30, 2002 this sequence version replaced gi:15290376.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES             Location/Qualifiers
     source          1. .192731
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-457H17"
BASE COUNT    54620 a  37176 c  40445 g  60490 t
ORIGIN


  Query Match             6.7%;  Score 78;  DB 9;  Length 192731;
  Best Local Similarity   46.3%;  Pred. No. 0.00039;
  Matches  327;  Conservative    0;  Mismatches  375;  Indels    4;  Gaps    2;

Qy      96 atatatatatgcatgcatgaggaccatggctatgatgaaggttaatagaggtagttgtga 155
           | ||||||||  ||   | |      |    |||  | ||   ||    |  || |
Db  163234 AAATATATATATATTTTTAATATATAATTATATATTAAAAAATATATAATATATTATAAT 163175

Qy     156 ttgagatatgtccagcactagttttttgttggtgtgatttctcatgatgacgcgaaaatt 215
           |  | ||      |    || | | || |   | |   ||   ||    |   | ||||
Db  163174 TATATATTAAAATATATATAATATATTATAATTATATATTAAAATATATATAATATAATT 163115

Qy     216 ttatatatatatataatgaataatatgattgattattctctgtaattttgtgaaatagat 275
           |||||  | |||||   ||||  || ||  ||||   | | | ||  | |   |||  |
Db  163114 ATATATTAAAATATATATAATATAATTATATATTAAAATATAT-ATAATATATTATAATT 163056

Qy     276 taaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggcatggtg 335
           |   |  |  || ||   | |      | | |||  |    |    |       || ||
Db  163055 ATATTAAAATATATATGTTATAATTGTATATTAAAATATATATAATATATTATAATTGTA 162996

Qy     336 atttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattat---ta 392
           || | | ||   ||   ||| |  | | |  |  || |  | || ||  ||||   ||
Db  162995 TATTATAATATATAAATATAATCTATATATTTTATATATATTATATATATTTTATATATA 162936

Qy     393 aattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtatt 452
           |  |||||  | |  |       |    ||   |||   || ||    | |  ||| |||
Db  162935 ATATATTACATAATATATATAATATATTACATAATATATAATATACCATATTATATTATA 162876

Qy     453 tttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaaaac 512
           |   ||| | ||  ||| || ||||  ||| | |   |    |  |      | || | |
Db  162875 TAATATATTATATATAATATATTATATGATATAATATATAATATATTATATATAATATAA 162816

Qy     513 ctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaat 572
           |   ||   | |'|||| ||||| | || | ||||| || ||  |  | || |
Db  162815 TATTATATTATATATTATAATGTAATATATTATATACATATATATTATACATATATATAC 162756

Qy     573 ctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaa 632
           |  |  |  | |  || | || |  ||  ||     || |||  | | |   || |  |
Db  162755 ATATATATAATATATTATATAATATATAATATAATAATATATTATATATCATATAATATA 162696

Qy     633 actggccaggacaattattaaaattaataattattatttttttctaataaagcacttccta 692
           |      | | |  | || ||| |  | |||  | | |  |   ||     |  ||
Db  162695 ATAATATATTATATATCATATAATATAATAATATATTATATAACATATAATATAATAATA 162636

Qy     693 attgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagta 752
           |  || |||||||     |  |   |  ||| ||| ||||     || | |   | ||
Db  162635 TATTATATAATATATAATATATTATATAATATATAATATATTATATAATATATAATATTA 162576

Qy     753 ggtgagaggtgctgacaaataaattagtgcataaaatataatggat 798
           | | |  |  |    ||||   ||    ||| |||||  ||  ||
Db  162575 TATAATATATAATATATAATATTATATATTATATAATATTATATAT 162530


Search completed: July 30, 2002, 11:28:25
Job time: 6859 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 10:05:46 ; Search time 209.08 Seconds
(without alignments)
9624.181 Million cell updates/sec

Title: US-09-530-663B-5
Perfect score: 1172
Sequence: 1 tgtaggattggtggaatggg.....agcccgcaatggacgcgaca 1172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1170.4	99.9	1172	20	AAX58644	Aspen 4-coumarate
c	2	72.2	6.2	9155	24	ABL32462	Human immune syste
	3	72.2	6.2	40862	24	ABL34073	Human immune syste
	4	71.8	6.1	14919	22	AAS46506	Tumour suppressor
c	5	70.2	6.0	2279	23	ABL18982	Drosophila melanog
	6	70.2	6.0	513445	22	AAI61373	Soybean 318013 reg
c	7	69.6	5.9	8323	24	ABL32058	Human immune syste
c	8	69	5.9	6071	24	ABL32325	Human immune syste
c	9	69	5.9	6071	24	AAS61076	Human gene regulat
c	10	68.8	5.9	21313	22	AAK82710	Human immune/haema
c	11	68.6	5.9	19787	24	ABL33451	Human immune syste
c	12	68.6	5.9	113515	24	ABL34175	Human immune syste
	13	67.4	5.8	10191	24	ABL34255	Human immune syste
c	14	67.4	5.8	17183	24	ABL32487	Human immune syste
	15	66.8	5.7	9706	22	AAK86270	Human immune/haema
	16	66.6	5.7	700	22	AAH93026	Human inflammatory
c	17	66.6	5.7	7498	24	ABL32257	Human immune syste
c	18	66.6	5.7	8305	24	ABL33569	Human immune syste
c	19	66.4	5.7	8964	24	ABL33534	Human immune syste
c	20	66.4	5.7	8964	24	AAS61276	Human gene regulat
c	21	65.6	5.6	19380	24	AAS61426	Human gene regulat
c	22	65.4	5.6	6109	24	ABL32326	Human immune syste
c	23	65.4	5.6	6109	24	AAS61077	Human gene regulat
	24	64.8	5.5	10328	24	ABL33545	Human immune syste
c	25	64.4	5.5	5398	24	ABL32142	Human immune syste
c	26	64.2	5.5	5798	24	ABL33667	Human immune syste
	27	64.2	5.5	18154	24	ABL32255	Human immune syste
	28	64	5.5	1132	21	AAF12929	Aspergillus oryzae
c	29	64	5.5	5311	24	ABL33019	Human immune syste
	30	64	5.5	6418	24	ABL32323	Human immune syste
	31	64	5.5	6418	24	AAS61074	Human gene regulat
	32	63.8	5.4	1915	20	AAX58642	Aspen 4-coumarate
	33	63.8	5.4	18624	24	ABL33703	Human immune syste
	34	63.8	5.4	19380	24	AAS61427	Human gene regulat
	35	63.4	5.4	4590	7	AAN60472	Sequence encoding
c	36	63.2	5.4	700	22	AAH93026	Human inflammatory
c	37	63.2	5.4	6130	22	AAS46646	Tumour suppressor
c	38	63.2	5.4	6130	24	ABL32129	Human immune syste
c	39	63.2	5.4	6361	24	ABL33141	Human immune syste
	40	63	5.4	9599	22	AAK86271	Human immune/haema
	41	62.8	5.4	5218	24	ABL33267	Human immune syste
c	42	62.6	5.3	14919	22	AAS46505	Tumour suppressor
c	43	62.6	5.3	15548	24	ABL34155	Human immune syste
	44	62.4	5.3	20674	21	AAC58017	Arachidonic acid m
	45	62.2	5.3	1864	15	AAQ78892	Aspergillus nidula

ALIGNMENTS

RESULT 1
AAX58644
ID AAX58644 standard; cDNA; 1172 BP.
XX
AC AAX58644;
XX
DT 16-AUG-1999 (first entry)
XX
DE Aspen 4-coumarate coenzyme A ligase Pt4CL1 gene promoter.
XX
KW Aspen; 4-coumarate coenzyme A ligase; lignin; transgenic plant;
KW tree; conifer; forestry; Pt4CL1; promoter; xylem; ss.
XX
OS Populus tremuloides.
XX
FH Key Location/Qualifiers
FT misc_feature 888..893
FT /*tag= a
FT /note= "box A, cis-acting element"
FT misc_feature 914..924
FT /*tag= b
FT /note= "box L, cis-acting element"
FT misc_feature 1018..1033
FT /*tag= c
FT /note= "box P, cis-acting element"
XX
PN WO9924561-A2.
XX
PD 20-MAY-1999.
XX
PF 12-NOV-1998; 98WO-US24138.
XX

```
PR   12-NOV-1997;   97US-0969046.
XX
PA   (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI   Chiang VLC,  Hu W,  Tsai C;
XX
DR   WPI; 1999-327394/27.
XX
PT   Altering properties of plants by modulating 4-coumarate co-enzyme A
PT   ligase
XX
PS   Disclosure; Page 71; 73pp; English.
XX
CC   This is the nucleotide sequence of the promoter region of the
CC   quaking aspen (Populus tremuloides Michx.) Pt4CL1 gene (see
CC   also AAX58642) that codes for 4-coumarate coenzyme A ligase 4CL1 (see
CC   AAY06092).  The promoter DNA was isolated from an aspen genomic
CC   library by screening with Pt4CL1 cDNA.  The promoter drives gene
CC   expression exclusively in xylem tissue.  It can be used to
CC   manipulate gene expression, and hence to engineer traits of
CC   interest, in the xylem tissue of target plants, e.g. to manipulate
CC   lignin content or structure, or to enhance growth, cellulose
CC   content or other value-added wood qualities.  Plants with altered
CC   contents of lignin and/or cellulose can be processed more
CC   efficiently, e.g. for pulp production, with reduced costs and
CC   pollution associated with lignin removal.
XX
SQ   Sequence 1172 BP; 399 A; 225 C; 180 G; 368 T; 0 other;


  Query Match             99.9%;  Score 1170.4;  DB 20;  Length 1172;
  Best Local Similarity   99.9%;  Pred. No. 8.5e-198;
  Matches 1171;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgtaggattggtggaatgggatcattcctaatcccttaatgacggtggcatgaacacaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tgtaggattggtggaatgggatcattcctaatcccttaatgacggtggcatgaacacaaa 60

Qy     61 gcaaagagaagttaggtcactcctcctttatatatatatatatatgcatgcatgaggacc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gcaaagagaagttaggtcactcctcctttatatatatatatatatgcatgcatgaggacc 120

Qy    121 atggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcactagtttt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 atggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcactagtttt 180

Qy    181 ttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataatgaataata 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataatgaataata 240

Qy    241 tgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgacc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgacc 300

Qy    301 agttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaatttg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 agttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaatttc 360

Qy    361 aaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggct 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 aaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggct 420

Qy    421 aatcaattattattaattaaaacgatagtatttttgataatttaattaaaattttattgg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aatcaattattattaattaaaacgatagtatttttgataatttaattaaaattttattgg 480

Qy    481 atttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatat 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 atttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatat 540

Qy    541 aatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 aatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgta 600

Qy    601 taattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 taattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaat 660

Qy    661 aattattatttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 aattattatttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaa 720

Qy    721 taataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 taataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagt 780

Qy    781 gcataaaatataatggattggtggtctgtgaaaagacaggtggaggacaagccacctctc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 gcataaaatataatggattggtggtctgtgaaaagacaggtggaggacaagccacctctc 840

Qy    841 tcaagtcaaaaggccatttcacaaccaacccaaatgggaacccaccaccgttccccgcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 tcaagtcaaaaggccatttcacaaccaacccaaatgggaacccaccaccgttccccgcca 900

Qy    901 ttaaaatccctaatctcaccaacccaactccacagattcttcaccaaacgcaactgattt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ttaaaatccctaatctcaccaacccaactccacagattcttcaccaaacgcaactgattt 960

Qy    961 ttcaatcaatgttttccctatactaccccccccaacaactccataatacccaatttgtcct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ttcaatcaatgttttccctatactaccccccccaacaactccataatacccaatttgtcct 1020

Qy   1021 ttcaccaaccccccgtcctccgtgccagccaattctatatcagcaggaatgctctgcactc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ttcaccaaccccccgtcctccgtgccagccaattctatatcagcaggaatgctctgcactc 1080

Qy   1081 tgctttctcaggtctcctaccataagaaaacagagagcacctaaaactcgccatctctcc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tgctttctcaggtctcctaccataagaaaacagagagcacctaaaactcgccatctctcc 1140

Qy   1141 ctctgcatctttagcccgcaatggacgcgaca 1172
          |||||||||||||||||||||||||||||||
Db   1141 ctctgcatctttagcccgcaatggacgcgaca 1172


RESULT   2
ABL32462/c
ID   ABL32462 standard; DNA; 9155 BP.
XX
AC   ABL32462;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 435.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
```

XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 435; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 9155 BP; 2719 A; 122 C; 1814 G; 4500 T; 0 other;

Query Match 6.2%; Score 72.2; DB 24; Length 9155;
Best Local Similarity 47.9%; Pred. No. 0.00037;
Matches 238; Conservative 0; Mismatches 258; Indels 1; Gaps 1;

```
Qy    210 aaaatttttatatatatatataatgaataatatga-ttgattattctctgtaattttgtga 268
          ||||    | |   | | | ||| | |||||| | || || | ||  |   | |   | |
Db   7842 AAAAAAAAAAAATAAAAAACAATTATTAATATCACTTCATAAATCAATAACAATAATTAA 7783

Qy    269 aatagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttgggg 328
          | | |  ||||||  |  ||   |   |  | |    ||  |||| |    ||
Db   7782 CAAATAAAAAAACAAATTTATAAAATTAAATTACCTAATTAAAAACCAATTATATTTTTT 7723

Qy    329 catggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatt 388
          |     |||||   || |  ||    ||   ||    |||  ||| ||| ||  | |
Db   7722 TAAAAATATTTTAATAACCTAAAAATTATACAAATCAAAAATATAACAAAAATCAACAAA 7663

Qy    389 attaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatag 448
          || |  |  |||   || |            || |  |   | | | |||||     ||
Db   7662 ATAATTTATTTACAATATTAATATATAAATATATTTTAAATTAAATTATTAATTATTTAC 7603

Qy    449 tatttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaa 508
          ||||||  |    |||   || |||    |   |   ||  |  | || ||| | ||||
Db   7602 TATTTTATAACTATTATAAAATATTAATATATCTAAAAAATATTTTAAATACTTAACAAT 7543

Qy    509 aaacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttt 568
          || ||||| ||||  | |  | |   | || |  | | |   ||||||  | |||
Db   7542 TAAACTAATAAAATACACAAATAAATCGCTAAATATAAAACTAATAAATTAATAAAAAAA 7483

Qy    569 aaatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattat 628
          |||  |  | |||   | ||    || || || |  | |   | |  ||| |  ||  |
Db   7482 TAATTTATAATTTTAATCATTTTTAATACCTAAATATCGAAAAAAAATAAAATTTAAAAA 7423

Qy    629 tcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcactt 688
            |  |  ||   | | |   ||  |    || || ||||      | | ||   ||||
Db   7422 AATATATTACCTTAAAATTCTATATTAAATTTATTTTTTATAAAACATTAAAACTAACTT 7363

Qy    689 cctaattgttaaaatat 705
          ||  ||  ||||| ||
Db   7362 ACTTTTTTATAAAACAT 7346
```

RESULT 3
ABL34073
ID ABL34073 standard; DNA; 40862 BP.
XX
AC ABL34073;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2046.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2046; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 40862 BP; 14301 A; 414 C; 7426 G; 18721 T; 0 other;

Query Match 6.2%; Score 72.2; DB 24; Length 40862;
Best Local Similarity 45.8%; Pred. No. 0.00039;
Matches 287; Conservative 0; Mismatches 338; Indels 2; Gaps 1;

```
Qy    177 ttttttgttggtgtgatttctcatgatgacgcgaaaatttatatatatatataatgaat 236
          ||| ||  || | | |  | ||  | | |    |||  |   |  | |||||  | |    |
Db  12813 tttattagtgtttttaaatgtcgaggtaaaataaaagtaaaagttttatattttaatttt 12872

Qy    237 aatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggt 296
          |    ||    |  | |  ||||  ||  ||||||   |        | | | | || |
Db  12873 tagtatatattattatgtaagtaagatttagaaataatattttggtatttatttttagtt 12932

Qy    297 gaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaa 356
          |    ||  |  ||  |  | |  |  |      ||  |    |   |  | | |    | |
Db  12933 gttatgtgtt--aaatataattgtaaattatttataatataataaaatttaaagttttat 12990

Qy    357 tttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggtt 416
          || || ||   ||||| |||    |||| ||  |   | | |    ||||   |  |
Db  12991 ttagataaaataaatttaaaggcgtttatatattattaatttgttttttaaaaaaaagaaa 13050

Qy    417 ggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatttta 476
          | |  | | ||  || | |     |   |  ||   || || ||||  || || | | |
Db  13051 tgtttttttatttggtagttaagtttatttttatagaatttattatttttttttaattgtaa 13110

Qy    477 ttggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtg 536
          | ||||   |   | | ||||   |   |    |   |  || |||| |   ||| |
Db  13111 tgggataatatgttaataaatttattataagttgaaagtaataagttaaaaatgtatttt 13170
```

```
Qy     537 atataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaaca 596
           ||||| || |  ||  ||  || ||   |       |  ||||| | ||| | | |
Db   13171 atatatttaatttatcgaagcgaatattatagtttagtatagtttatttttaaatatattt 13230

Qy     597 cgtataattgggctagatttaacagctattattcaaactggccaggacaattattaaaat 656
           |||  |||  |   |  ||||      | ||    || |    |  |   |||||  |
Db   13231 agtaatatttatattagtttatagttgagtaaaataatttaatataaagtttatttttata 13290

Qy     657 taataattattatttttttctaataaagcacttcctaattgttaaaatatatgtctaaaca 716
           | |  | ||||||||  | || | ||  | |       | || |||   | |  || |
Db   13291 aagtgttgattattttatgtatttaataaataatattgtattgaaagtgaagaatagaag 13350

Qy     717 ctaataataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaat 776
           |   |  | |  || | | |   |||   | ||      || ||   |    ||| |
Db   13351 tatagtttttattgaatgtttattgtttttatattatcgtgaagttgaaaaattttaagt 13410

Qy     777 tagtgcataaaatataatggattggtg 803
           ||    || | |  | | ||||| |||
Db   13411 taaattattataagttacggatttgtg 13437
```

```
RESULT   4
AAS46506
ID   AAS46506 standard; DNA; 14919 BP.
XX
AC   AAS46506;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #228.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 228; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 14919 BP; 4881 A; 249 C; 2861 G; 6926 T; 2 other;
```

```
  Query Match             6.1%;  Score 71.8;  DB 22;  Length 14919;
  Best Local Similarity   45.2%;  Pred. No. 0.00044;
  Matches  344;  Conservative    0;  Mismatches  412;  Indels    5;  Gaps    2;

Qy      58 aaagcaaagagaagttaggtcactcctcctttatatatatatatatatgcatgcatgagg 117
           |||   ||   ||| ||| | | |     ||   | || |  ||  ||   |  ||
Db    7239 aaaagtaacgaaaggtagtttaatataagttgtgaaattttaattaatttgttaatttat 7298

Qy     118 accatggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcactagt 177
           |     ||   |||   || || |  ||     |       |||| |  |    ||
Db    7299 ttaaataaaatttagaattatattatatatatgattttaaattatattttaatttatata 7358

Qy     178 tttttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataatgaata 237
           |||||||  |  ||||      | |||        || |||||  |||     ||  |
Db    7359 attttgtttataagattatgatttatgttatttttatattatagttattatatatttaat 7418

Qy     238 atatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtg 297
           |   |  || | | | || ||  | || | || || | |  |   |     |
Db    7419 tatatatatttatttatatatgatatttggtaagaaatgaagaaatgtttttaaatattt 7478

Qy     298 accagttgtcaaa---tgaccactcgacttggggcatggtgatttttcaaatcacaactc 354
           | |    |||   |     |    ||     ||  |  | ||  ||||   |  |
Db    7479 cgtaatgaataaaagttatttttgaggttttaatattatttttattataaatagaattta 7538

Qy     355 aatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacggg 414
           |||| || ||  |||| | ||  |  |||   || ||   || || ||||| ||    ||
Db    7539 aattagataatgaaaaattaagtaatttttttttttttttttttttggttatttttcgg 7598

Qy     415 ttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaattt 474
           ||   | ||   ||   | ||    |     |   | ||||  || |  | ||    |||
Db    7599 ttaaattattgttttaaaatagaataggaatttagaatttttattagtgaatttttttttt 7658

Qy     475 tattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatg 534
           | ||  |  |   |  | |  ||||    | |  | |   | |    | ||||| || |
Db    7659 tttttaaggttttttataatagattaaggaaaagtatatagatttttgtaaatatttttttt 7718

Qy     535 tgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaa 594
           |  | || |||||   |   | |  || |    |  |  |  | |||   | ||     |
Db    7719 ttttttatttttaggtttggtagttttttaataattgaattatataatttaaataatttgtta 7778

Qy     595 cacgtataattgggctagatttaacagctattattcaaactggccaggacaattattaaa 654
           |  |  ||| |   ||  || |     | ||    |||      |  | ||  |  |||
Db    7779 tattttaaatagttttattttaatgtaatttttagaaaagattaaaaaaaaaaaaaaaaa 7838

Qy     655 a--ttaataattattatttttttctaataaagcacttcctaattgttaaaatatatgtcta 712
           |  | | || || ||| ||||  | |     | ||  | ||| ||  ||    | | |
Db    7839 agttaattatttttttaatttttaaatttttaaaatttttttattttttttaaatatttttttt 7898

Qy     713 aacactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaat 772
           | |  ||   || ||| ||| | |||  | |     |   || | |  |||  ||
Db    7899 tttatttttagggaatttttttttttttataatagtataaaatatagtgttaatgaataaa 7958

Qy     773 aaattagtgcataaaatataatggattggtggtctgtgaaa 813
           | || || |||| | |    || ||||  | |  ||
```

```
Db   7959 tttaaattgtattaaattttaaaattttgtgggttttataa 7999


RESULT   5
ABL18982/c
ID   ABL18982 standard; DNA; 2279 BP.
XX
AC   ABL18982;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 8419.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 8419; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2279 BP; 855 A; 260 C; 235 G; 929 T; 0 other;


  Query Match             6.0%;  Score 70.2;  DB 23;  Length 2279;
  Best Local Similarity   49.9%;  Pred. No. 0.00079;
  Matches  256;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    3;

Qy    210 aaaatttttatatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
          ||||| | |  || | | || ||  |||||     | ||||||  |     |  |
Db    515 AAAATCTCAACTAAAGAGATTATTTATAATTATTATTATTATTATTATAATATAATAATT 456

Qy    270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
          ||| |||| ||    | |||   |  | |  | || | || |    | |  |  |
Db    455 ATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAA 396

Qy    330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
          ||  | | || | || |   || |  |  | | || | | |    | ||  | || ||||
Db    395 ATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTA 336

Qy    390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
          | || |    |  ||| | |   ||   ||| |  ||||| ||||| ||  ||  |
Db    335 TAAATTATAAATTATAAATTATAAATTATA-AATTATAAATTATAAATTATAA--ATTAT 279

Qy    450 atttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaa 509
          |  ||  | | | ||| | | |  ||||    | | ||| |       || || | |||
Db    278 AAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAAT 219

Qy    510 aacctaatcaaattaatatcttatgtgatataatttagaaa-tataaatgattaacctttt 568
          |   | |  || | |||  ||||    ||||| ||| ||| |||||||| || ||   |
Db    218 TATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATA 159

Qy    569 aaatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattat 628
          || | |  | | |   |||||||  | |  |   |       |  | |||    ||   |
Db    158 AATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATT 99

Qy    629 tcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcactt 688
            ||| |    |  | || |  |||| |  | | |    ||| |   | |||||  |
Db     98 ATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAAATTATAA 39

Qy    689 cctaattgttaaaatatatgtctaaacactaat 721
            ||    ||| ||  |||   | | || | ||
Db     38 ATTATAAATTATAAATTATAAATTATCAATTAT 6


RESULT   6
AAI61373
ID   AAI61373 standard; DNA; 513445 BP.
XX
AC   AAI61373;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 318013 region A3, SEQ ID NO: 4.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US00552.
XX
PR   07-JAN-2000; 2000US-0174880.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
DR   P-PSDB; AAM42216.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs -
XX
PS   Claim 30; Page 596-893; 1353pp; English.
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule
CC   provided in the specification.
XX
SQ   Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;
```

```
  Query Match             6.0%;  Score 70.2;  DB 22;  Length 513445;
  Best Local Similarity   47.9%;  Pred. No. 0.00096;
  Matches  263;  Conservative    0;  Mismatches  283;  Indels    3;  Gaps    2;

Qy    210 aaaatttttatatatatatatatgaataatatgattgattattctctgtaattttgtgaa 269
          | |  | ||||||||||||||||   | | |||||||  ||||   |   |||  |     ||
Db 464592 atatatatatatatatatatatatattcatatgatgtattacatttattaaatcaaataa 464651

Qy    270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
          || |||| |         ||   |  | |  |  |  | |||    | |  |  |
Db 464652 gtatattatatggtaaatataatatttaataatattaccaattttaattgtatattctat 464711

Qy    330 atggtgatttttcaaatcacaactcaa-tttgaaaactaaaattaaaaaagatttagatt 388
          |  | |   |    |   | | |||| |||||| ||  |   |||   |  |||  | |
Db 464712 ttaataaaaataattaatgctagtcaattttgaatacggattataagctaatttttaaat 464771

Qy    389 attaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatag 448
          | |||   | ||  ||   |      ||   ||||  |||||| |  ||   ||   |
Db 464772 aataataaaataatttttatataaaattaattaatatattattttagatattaaacttt 464831

Qy    449 tatttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaa 508
          || |||  | |||||| || |    || ||   | |  ||  | |||  ||    | |
Db 464832 aatatttttttcatttagtttattgatttttttactatacatagattca--tatgatatatt 464889

Qy    509 aaacctaatcaaaattaatatcttatgtgatataatttagaaatataaatgattaaccttt 568
          | |   | | ||  |||||  |    |    | | | |  |  | | |  | || |   | |
Db 464890 acattcattaaatctaataaatacactatatttagtaaatataaaatttaataatattat 464949

Qy    569 aaatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattat 628
          |||  |   | | |  ||| | | || |  ||| | |  |  |||   ||   || | |
Db 464950 taatgttatttatttattaaatacaataattattactatgtaagaaaaaatcactttcaa 465009

Qy    629 tcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcactt 688
          |   |    ||  |||| ||||    |||||  ||   |  |  |   | |    |
Db 465010 atatgttttattggctaatttttaagtaaaataaaaatagctaatattttaatatatata 465069

Qy    689 cctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggc 748
            ||  | |  | |||||| | || | |   |||   | ||| || | |  |  || | |
Db 465070 tatatatatatatatatatatatatatatatatatatatattcatatattaactttggtt 465129

Qy    749 agtaggtga 757
          |  || | |
Db 465130 aaaagataa 465138


RESULT   7
ABL32058/c
ID   ABL32058 standard; DNA; 8323 BP.
XX
AC   ABL32058;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 31.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 31; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 8323 BP; 2499 A; 44 C; 1604 G; 4176 T; 0 other;


  Query Match             5.9%;  Score 69.6;  DB 24;  Length 8323;
  Best Local Similarity   47.4%;  Pred. No. 0.0011;
  Matches  240;  Conservative    0;  Mismatches  264;  Indels    2;  Gaps    1;

Qy    235 ataatatgattgattattctctgtaatttttgtgaaatagattaaaacagctcaatgtgag 294
          |||| |  | | |  ||| | |  | | || |  | || | | ||| |  |   |   |
Db   5598 ATAAAACCAATAAACATTTTATAAACTATTTTATATTAAAATTAAAAAAATATTTCAAAA 5539

Qy    295 gtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactc 354
              | |  |   |       ||    ||||    ||   | |  ||||  ||    ||
Db   5538 TCTTCTAAATCAAATCCTCTCATAAAACTTTCATCAAACTTAACTTTCTTATATTCACAA 5479

Qy    355 aatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacggg 414
          ||  |  |||   |||| | || | ||||| ||||| | |   | |  | ||   |
Db   5478 AAAATATAAATACAAATAATAATACATTTAAATTATAATACCTTAATATAAAAAAAACTT 5419

Qy    415 ttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaattt 474
            |  |||  ||   || | | || ||        |||  | |||||| | |||||||
Db   5418 CTTAATAAAAAACATTTCTAACTTTAACTTTACAAATTAATCATAATTAACTTAAAATAA 5359

Qy    475 tattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatg 534
            |     || |  | |   |  |  |  | ||||||  ||| ||  | |  |    |
Db   5358 ATTATCGCTTTACCGCAAATATCTTTAATATAAAAAA--TAAACACTTCATAAATACAAA 5301

Qy    535 tgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaa 594
          |  ||||  ||      || |||| ||||   | |  | |      | |  |  || | |
Db   5300 TTCTATACATTCTTTTTAAAAATATTTAAAAATAACTTATTTCCACTATAATTCAATATA 5241

Qy    595 cacgtataattgggctagatttaacagctattattcaaactggccaggacaattattaaa 654
          ||      |||     | || | |||  |  |||| ||     ||   | | | | ||
Db   5240 CATTACATATTATAACATATATTACATATCATATTTCAATATATCATTTCTAATTTAAAT 5181

Qy    655 attaataattattatttttttctaataaagcacttcctaattgttaaaaatatatgtctaaa 714
          || |||  | | || ||  || ||||    | | | |  ||  | | ||  ||    | |
Db   5180 ATAAATTTTAAATAATTCATCAAATACTATATTACTTTCTTTATTATATTAATAAAAATA 5121

Qy    715 cactaataataaaatttatttgtgta 740
            | |   || || | |   || | ||
Db   5120 TAATTCCAACAATAATATATTTTTTA 5095


RESULT   8
ABL32325/c
ID   ABL32325 standard; DNA; 6071 BP.
XX
AC   ABL32325;
XX
```

```
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 298.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 298; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6071 BP; 1973 A; 46 C; 1013 G; 3039 T; 0 other;


  Query Match             5.9%;  Score 69;  DB 24;  Length 6071;
  Best Local Similarity   48.4%;  Pred. No. 0.0013;
  Matches  256;  Conservative    0;  Mismatches  265;  Indels    8;  Gaps    2;

Qy     348 acaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaat 407
           | ||||| || | ||||  | ||  ||| || |  |    ||  |||||||||   ||  |
Db     651 AAAACTCCATCTCAAAAAAACAAACAAACAAAAACTCTTATAAAAAATTATTTTCTTTTT 592

Qy     408 tcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaatt 467
             ||       |  | |     | |   ||||||  | |   | |     ||   ||
Db     591 AAACAAAAAAACCTAACTTCATTAACATTTTAAAATAAAAACTTCTAAACAAAAAAAAAA 532

Qy     468 aaaattttattggatttgaatgaactcaattacatc--acaaaaaacctaatcaaattaa 525
               |  |     ||| ||  || |  | |  |||  |  |||||   ||  ||||  |
Db     531 TTTCATAAAAATATTTTAAAATAAATATAATCAATCTAATCAAAAAATAAAAAAAATCTA 472

Qy     526 tatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctct 585
             |  |||| |  ||||| | | | | | |||     ||  |||||||  |  | | | |
Db     471 AACTTTATCTCTTATAAATAAAATAAAAAAAAACAAAAAATTTAAATTTTAATTATATAA 412

Qy     586 tataaaaaacacgtataattgggctagatttaacagctattattcaaactggccaggaca 645
           |||||||||  |  ||   ||     | ||||||  |   |  | | |||||   ||  |
Db     411 TATAAAAAATAATTAATCTTAAAACAAATTTAAAATAAAAAAATTAAACTAATCAATATT 352

Qy     646 attattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaatat 705
```

```
           ||  ||   |||| || ||| |      |||||||| |  ||    | | |||||| |
Db     351 TTTTATACTTTTAAAAAATATAA------CTAATAAAACGTTTTTATACTTTTAAAAAAA 298

Qy     706 atgtctaaacactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgct 765
           |  |  | | |  ||   ||| | |   |    ||| ||   || |       |     |
Db     297 ACCTAAATAAATAAACTTTAATACTATATATATTATATTAAACATTTAAAATAACAAAAT 238

Qy     766 gacaaataaattagtgcataaaatataatggattggtggtctgtgaaaagacaggtggag 825
           |    |||| || | |  | ||  | ||       |  |||   |||  | |     |
Db     237 AATTTCTAAAATATTTCTAACAACTTCATTTTCATTTTATCTACAAAATAAAAAAATAAA 178

Qy     826 gacaagccacctctctcaagtcaaaaggccatttcacaaccaacccaaa 874
             ||       |   | || | || |  |||| | |  |   || ||||
Db     177 TTTAAAAATTATTAATTAAATTAACAAACCATATAATTATTTACGCAAA 129


RESULT   9
AAS61076/c
ID   AAS61076 standard; DNA; 6071 BP.
XX
AC   AAS61076;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human gene regulation-associated gene oligonucleotide #31.
XX
KW   Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW   cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW   asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW   renal disease; Preeclampsia; cardiac allograft vascular disease;
KW   colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW   immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW   nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200177375-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03968.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017470/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated
PT   with gene regulation, useful for analysing cytosine methylations for
PT   diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT   disease  -
XX
PS   Claim 1; SEQ ID No 32; 26pp; English.
XX
CC   The invention relates to 224 nucleic acid sequences comprising at least
CC   18 bases of a chemically pretreated gene associated with gene regulation
CC   selected from 43 known genes (or complementary sequences). The
CC   chemical pretreatment converts cytosine bases unmethylated at the
CC   5-position to uracil or another base with hybridisation behaviour
CC   dissimilar to cytosine, to enable analysis of cytosine methylations.
CC   The DNA sequences, oligomers (or sets/arrays) and method are
CC   useful in the diagnosis of diseases (or predisposition to diseases)
CC   associated with gene regulation and in therapy of such diseases, by
CC   enabling analysis of the cytosine methylation patterns of such genes,
CC   kits are provided. They are especially useful in diagnosis
CC   and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC   disorders, haemophilia, solid tumours and cancer, Werner syndrome,
```

```
CC   asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC   preeclampsia, graft versus-host disease. The present sequence is a
CC   sequence included in the sequence data for this specification and is
CC   associated with the human gene regulation-associated genes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6071 BP; 1973 A; 46 C; 1013 G; 3039 T; 0 other;


  Query Match             5.9%;  Score 69;  DB 24;  Length 6071;
  Best Local Similarity   48.4%;  Pred. No. 0.0013;
  Matches  256;  Conservative    0;  Mismatches  265;  Indels    8;  Gaps    2;

Qy    348 acaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaat 407
          | ||||| || | |||| | ||  ||| || | |    ||  |||||||| ||  |
Db    651 AAAACTCCATCTCAAAAAAACAAACAAACAAAAACTCTTATAAAAAATTATTTTCTTTTT 592

Qy    408 tcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaatt 467
            ||     | | | |     | |   ||||||  | |   | |    ||   ||
Db    591 AAACAAAAAAACCTAACTTCATTAACATTTTAAAATAAAAACTTCTAAACAAAAAAAAA 532

Qy    468 aaaattttattggatttgaatgaactcaattacatc--acaaaaaacctaatcaaattaa 525
              |  |    ||| ||  || |  | |  |||  |  |||||   ||  ||||  |
Db    531 TTTCATAAAAATATTTTAAAATAAATATAATCAATCTAATCAAAAAATAAAAAAAATCTA 472

Qy    526 tatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctct 585
            |  |||| |  |||| | | | | | | |||     ||  |||||| |  | | | |
Db    471 AACTTTATCTCTTATAAATAAAATAAAAAAAAACAAAAAATTTAAATTTTAATTATATAA 412

Qy    586 tataaaaaacacgtataattgggctagatttaacagctattattcaaactggccaggaca 645
          |||||||||  |  ||   ||     | |||||| |   |  | | ||||   ||  |
Db    411 TATAAAAAATAATTAATCTTAAAACAAATTTAAAATAAAAAAATTAAACTAATCAATATT 352

Qy    646 attattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaatat 705
            ||  ||  |||| || ||| |      ||||||||  |  ||    | | ||||||  |
Db    351 TTTTATACTTTTAAAAAATATAA------CTAATAAAACGTTTTTATACTTTTAAAAAAA 298

Qy    706 atgtctaaacactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgct 765
          |  |  | | |  ||   ||| | |   |    ||| ||   || |       |     |
Db    297 ACCTAAATAAATAAACTTTAATACTATATATATTATATTAAACATTTAAAATAACAAAAT 238

Qy    766 gacaaataaattagtgcataaaatataatggattggtggtctgtgaaaagacaggtggag 825
          |    |||| || | |  | ||  | ||       |  |||   |||  | |       |
Db    237 AATTTCTAAAATATTTCTAACAACTTCATTTTCATTTTATCTACAAAATAAAAAAATAAA 178

Qy    826 gacaagccacctctctcaagtcaaaaggccatttcacaaccaacccaaa 874
            ||      |   | || | || |  |||| | |  |   || ||||
Db    177 TTTAAAAATTATTAATTAAATTAACAAACCATATAATTATTTACGCAAA 129


RESULT  10
AAK82710/c
ID   AAK82710 standard; DNA; 21313 BP.
XX
AC   AAK82710;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37522.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.

XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
```

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 37522; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 21313 BP; 5220 A; 5786 C; 5273 G; 5034 T; 0 other;

Query Match 5.9%; Score 68.8; DB 22; Length 21313;
Best Local Similarity 48.2%; Pred. No. 0.0015;
Matches 193; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```
Qy    400 aggttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgata 459
          |||   |||   ||| |    |   |   | | |  || |||||     ||| | | |||
Db  17149 AGGGCTATTATGGGGGTTATAAGCTAGGGACTGTGGATGAAAACCGATATATATATTATA 17090

Qy    460 atttaattaaaattttattggatttgaatgaactcaattacatcacaaaaaacctaatca 519
            |||  |||   | ||||  || | |||  |    ||    |    ||  |  || |
Db  17089 TATTATATAATTATATATTATATATAAATATATATTATATATTATATAATTATATATTAT 17030

Qy    520 aattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagt 579
          |  |||||| | ||   ||||||| || ||     ||| ||  |   |  | | |  | |
Db  17029 ATATAATATATAATTACATATAATATATAATATATAATTATATATTATATATTATATATT 16970

Qy    580 ttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggcc 639
           | | | | | | |  |  |||||||       || ||  |  ||| |||  |  |
Db  16969 ATATATAAAATATATAATATATAATTATATATTATATATAATATATAATTATATATAATA 16910

Qy    640 aggacaattattaaaattaataattattatttttttctaataaagcacttcctaattgtta 699
            | |||||   | | ||   | |  ||| || | ||  |    |  |  ||  |  ||
Db  16909 TATATAATTAGATATAATATATAATTATATATTATATATAATTATATATAATATATTATA 16850

Qy    700 aaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggtgaga 759
          |  ||       || ||  | |   | ||     | | |||  |    | ||  |   |
Db  16849 TATAATTATATATAATACATAAATATATATAATAGTATATATTATATATAATATATATAA 16790

Qy    760 ggtgctgacaaataaattagtgcataaaatataatggatt 799
            |    | | ||| | ||||  || | |||||||  |||
Db  16789 TATATAAATATATATAATAGTATATTATATATAATATATT 16750
```

RESULT 11
ABL33451/c
ID ABL33451 standard; DNA; 19787 BP.
XX
AC ABL33451;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1424.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

```
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1424; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 19787 BP; 6302 A; 270 C; 3756 G; 9459 T; 0 other;


  Query Match             5.9%;  Score 68.6;  DB 24;  Length 19787;
  Best Local Similarity   51.9%;  Pred. No. 0.0016;
  Matches  203;  Conservative    0;  Mismatches  184;  Indels    4;  Gaps    2;

Qy     336 atttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaat 395
           |||||    |  | |||| | |||   || |||||   | ||| ||||  |  | ||  |
Db    2490 ATTTTAATCACAAAAACTTATTTTACTAAATAAAAAATACAAAAATTTTAAAAACTATCT 2431

Qy     396 tattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttt 455
           | | |   ||||| |    ||      |  ||| | |   || |  ||  | | |||||
Db    2430 TCTAATTATAATTTAACATTTCAACTAAAAATAACTTAAAAATTCAATAAAACTATTTCA 2371

Qy     456 gataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaaaaccta 515
           ||   | |||||||  |        |  | ||   |   || |  ||     || | ||
Db    2370 ACTACAATTATTAAAAATCATAATAAACTAAAATTAAATAACTTTATTCTTCAATATCTT 2311

Qy     516 atcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctc 575
             |||| |  || ||  || ||  ||  |  ||| |   || | ||||  ||    |||
Db    2310 CCAAAATAACGATATTTCGTAATTAAAAATTAAAAAACTCATTA-TAACTATTTTTCCTC 2252

Qy     576 gagtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaact 635
             ||| || |||| || |||| || || |   |||   ||  |   ||   || |||
Db    2251 CTATTTTTCATATATAATACACTTAAAAATAACCTA---TTCTCTTATACATTTAAAATA 2195

Qy     636 ggccaggacaattattaaaattaataattattatttttttctaataaagcacttcctaatt 695
             |    |||| || ||||| |    ||  | || ||| |   |||  |  || ||||
Db    2194 TTCATTATCAATAATAAAAATAATATTTTTATTTTATTTTTCTCAAACAAAATCTTAATC 2135

Qy     696 gttaaaatatatgtctaaacactaataataa 726
           | |   |||     | ||| ||||  |||
Db    2134 TATTACCCATACTAAAATACAATAATCTTAA 2104


RESULT  12
```

```
ABL34175/c
ID   ABL34175 standard; DNA; 113515 BP.
XX
AC   ABL34175;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2148.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2148; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 113515 BP; 35172 A; 1174 C; 22520 G; 54649 T; 0 other;


  Query Match             5.9%;  Score 68.6;  DB 24;  Length 113515;
  Best Local Similarity   48.8%;  Pred. No. 0.0017;
  Matches  185;  Conservative    0;  Mismatches  194;  Indels    0;  Gaps    0;

Qy     210 aaaatttatatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
           ||||  | | | |||||||||   ||    |  ||| | ||   | | |||   | | ||
Db   25148 AAAAAATAAAAAATATATATATCAAAACTAACTATTTAATAACATATATAAACCTTTTAA 25089

Qy     270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
           | |      ||| |  | | |   |    |  | |  | ||      |    | ||
Db   25088 AAATTCATAAAAACATTATTAAAATAATATTACTATTTAATAATAAATATAATTTCTATA 25029

Qy     330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
           ||  |  ||||  |||  | ||  ||| |  ||||    ||  |  ||        |||
Db   25028 ATATTACTTTTATAAACAATAATCCAAATATAAAAAATTAAAAATTAACACAAACAATTT 24969

Qy     390 ttaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatagt 449
           |||| |  | |  |||| |       |   |||  |||||| |||| || |||| | | |
Db   24968 TTAACTCCTCAACTTAAATACAACAATTTTTAAATAATTATCATTATTTTAAACTAAATT 24909
```

```
Qy    450 atttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaa 509
Db  24908 TATAACAAATAATTATTTAAATATTTATTACATTTTTATAACATTTACTTCATCCAATTA 24849

Qy    510 aacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttta 569
Db  24848 AATTTATTTCCTCAATTACACAATAATTTCCATATAATTATTCTTTTTTTTTTTTTTTT 24789

Qy    570 aatctcgagtttctcttat 588
Db  24788 TAACGAAATTTTACCTTAT 24770
```

RESULT 13
ABL34255
ID ABL34255 standard; DNA; 10191 BP.
XX
AC ABL34255;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2228.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2228; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 10191 BP; 2964 A; 285 C; 2323 G; 4617 T; 2 other;

Query Match 5.8%; Score 67.4; DB 24; Length 10191;
Best Local Similarity 46.2%; Pred. No. 0.0026;
Matches 224; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

```
Qy     87 tttatatatatatatatatgcatgcatgaggaccatggctatgatgaaggttaatagagg 146
Db   3592 tttgttattatatattttgggaatatttttaatagtaaatattgtattttaaattaagat 3651

Qy    147 tagttgtgattgagatatgtccagcactagtttttttgttggtgtgatttctcatgatgac 206
Db   3652 aattattattttggatttgtagaataatagatgtttagtattttagtatttaaaaatata 3711

Qy    207 gcgaaaatttttatatatatatataatgaataatatgattgattattctctgtaattttgt 266
Db   3712 ataaagattagttttatagttttataaataattatttttaattgtaaaataatagtttat 3771

Qy    267 gaaatagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttgg 326
Db   3772 ttttttttataataatatgaaaatttgtattattaattgtttaaatggttagtttgttttg 3831

Qy    327 ggcatggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttaga 386
Db   3832 ttgatttcgaaagttgatgtttattaaatttttgttatgaaaaaatggagtatttcggtta 3891

Qy    387 ttattaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgat 446
Db   3892 ttagtgaaatattgtatgtattttaggtttcgtttattaaataagttttaaaaatagaat 3951

Qy    447 agtatttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcaca 506
Db   3952 attttatttttaaatgtttttaaaaaatgtttttaatttgtaaagttttaagtttgttttt 4011

Qy    507 aaaaacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaacct 566
Db   4012 tttaaattaggtagattatttttatatttttttttagtatttttaatttaaggatggagat 4071

Qy    567 ttaaa 571
Db   4072 ttata 4076
```

RESULT 14
ABL32487/c
ID ABL32487 standard; DNA; 17183 BP.
XX
AC ABL32487;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 460.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal

```
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 460; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 17183 BP; 5342 A; 149 C; 3418 G; 8274 T; 0 other;


  Query Match             5.8%;  Score 67.4;  DB 24;  Length 17183;
  Best Local Similarity   47.7%;  Pred. No. 0.0027;
  Matches  197;  Conservative    0;  Mismatches  216;  Indels    0;  Gaps    0;

Qy     192 atttctcatgatgacgcgaaaattttatatatatatataatgaataatatgattgattat 251
           | || | || ||      | |  | |||| ||||||||    || || |  ||  |  ||
Db   12845 AATTATAATTATATAATCACATATATATACATATATATCTATAAAAAAAAAATAAAAAAT 12786

Qy     252 tctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgaccagttgtcaaat 311
             |    ||  |     || | ||||| | |    ||    |  | || | |    |||
Db   12785 AATAAAAAAAATACAACAAAATATTAACATACAAAAAATCTAAATAACAAATATATAAAA 12726

Qy     312 gaccactcgacttggggcatggtgatttttcaaatcacaactcaatttgaaaactaaaat 371
             | ||  | |       |      || |  |||    |  ||||  | ||||  ||||
Db   12725 ATTCTCTACAATACTCTTAAAACTTTTCTATAAAAATTACTTCAAAATAAAAAACAAAAC 12666

Qy     372 taaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaattatt 431
            ||||  |  |     |  | || |   ||  | || | ||        |     | | |
Db   12665 AAAAATAACTACTATTCCTAAATTCCCTAAATAAACTTACACTACTTTTTCAAACTAAAA 12606

Qy     432 attaattaaaacgatagtatttttgataatttaattaaaattttattggatttgaatgaa 491
           | ||| |  |   |||   |||   |||| || | |||| || || |  |  | ||  ||
Db   12605 AATAAATCCACTAATAAATTTTAAAATAAATTTAATAAATTTCTAATAAAAATAAAATAA 12546

Qy     492 ctcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataatttagaaat 551
           |  ||    |||  | | |||  ||   ||   | |          |||||  |||  ||
Db   12545 CCTAACCTAATCTAACATAACACAACATAACACAAAAAACGAACATTATAAATTACTAAA 12486

Qy     552 ataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataat 604
           |  ||  | |||| | ||||  |    |||   || ||| ||| |     |||
Db   12485 AATAAATACTAACTTATAAAAATTTTATTTAAATTCTAATAAAAAAAACAAAT 12433


RESULT  15
AAK86270
ID   AAK86270 standard; DNA; 9706 BP.
XX
AC   AAK86270;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41082.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
```

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 41082; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 9706 BP; 3114 A; 1643 C; 1682 G; 3267 T; 0 other;

Query Match 5.7%; Score 66.8; DB 22; Length 9706;
Best Local Similarity 50.3%; Pred. No. 0.0033;
Matches 190; Conservative 0; Mismatches 187; Indels 1; Gaps 1;

```
Qy    351 actcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattca 410
          | |  || | | |  ||| ||  || |   | || |||||  |     ||   ||| | |
Db   7540 aatatatattatatataatatataatatattatatattatatataatatattataaatta 7599

Qy    411 cgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaa 470
            ||   | || |  ||| ||  |||| |       |||  |  ||||| || |  |
Db   7600 tatattatat-ataatatataatatattatatattatatataatatataatatattatat 7658

Qy    471 attttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatct 530
          ||| |||   || |     || | |  ||  | | | || |  | || || |  |||| |
Db   7659 attatatataatatattataaattatatatttttatataatatattataaattatatattt 7718

Qy    531 tatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataa 590
          ||| | |||||| ||| | || | |   |||| | |||||     |||| | | |||
Db   7719 tatataatataagttatatattttatatattatattataaattatatgtttatataatat 7778

Qy    591 aaaacacgtataattgggctagatttaacagctattattcaaactggccaggacaattat 650
          |  | |  |   ||       |   ||  |  |||  || | |      |  | || |
Db   7779 attataaattatatatttttatgtaatataaattatatttttatataatatattataaatta 7838

Qy    651 taaaattaataattattatttttctaataaagcacttcctaattgttaaaatatatgtc 710
          |||| |  ||| |  |||| |   | ||| |  |  |  ||  | ||||      | |
Db   7839 taaattatatattatatattataaattatatattatataatatatattaattatattttt 7898

Qy    711 taaacactaataataaaa 728
            | | |   |||   |||
Db   7899 catatatatatatataaa 7916
```

Search completed: July 30, 2002, 11:22:21
Job time: 4595 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 09:06:51 ; Search time 1681.15 Seconds
(without alignments)
9409.292 Million cell updates/sec

Title: US-09-530-663B-5
Perfect score: 1172
Sequence: 1 tgtaggattggtggaatggg.....agcccgcaatggacgcgaca 1172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	99	8.4	1101	12	CNS0039G	AL063921 Drosophil
2	94	8.0	1101	12	CNS00EVL	AL069706 Drosophil
c 3	93.4	8.0	1101	12	CNS00EVL	AL069706 Drosophil
4	91.4	7.8	1101	12	CNS00EO7	AL069440 Drosophil
5	89.2	7.6	1101	12	CNS0039G	AL063921 Drosophil
c 6	84.8	7.2	1101	12	CNS00EO7	AL069440 Drosophil
7	84.8	7.2	1190	12	CNS02ON7	AL206908 Tetraodon
8	84	7.2	1101	12	CNS0021J	AL061936 Drosophil
9	82.2	7.0	1101	12	CNS003BD	AL064091 Drosophil
10	81.2	6.9	910	12	CNS01G8P	AL142826 Anopheles
c 11	80.6	6.9	734	12	CNS010MP	AL099163 Drosophil
12	79.6	6.8	1092	12	CNS020K7	AL175696 Tetraodon
13	79	6.7	1101	12	CNS003BB	AL064089 Drosophil
14	78.4	6.7	945	12	CNS04D0K	AL285149 Tetraodon
c 15	78.4	6.7	1101	12	CNS0039L	AL063926 Drosophil
16	78	6.7	1013	12	CNS06RPQ	AL412260 T7 end of
c 17	77.6	6.6	1201	12	CNS0167M	AL106396 Drosophil
18	77	6.6	1101	12	CNS00EQL	AL069526 Drosophil
19	77	6.6	1143	9	AL565457	AL565457 AL565457
c 20	77	6.6	1143	9	AL565457	AL565457 AL565457
c 21	76.2	6.5	1101	12	CNS001FB	AL060732 Drosophil
c 22	76	6.5	928	12	CNS00DKY	AL071865 Drosophil
23	76	6.5	1225	12	CNS0161D	AL106171 Drosophil
c 24	75.8	6.5	1092	12	CNS020K7	AL175696 Tetraodon
25	75.8	6.5	1101	12	CNS00BO1	AL057419 Drosophil
c 26	75.6	6.5	886	12	BH177277	BH177277 008_L_22-
c 27	75.6	6.5	886	12	CNS07JUX	AL614235 T3 end of
28	75.4	6.4	734	12	CNS010MP	AL099163 Drosophil
c 29	75.2	6.4	1101	12	CNS00EQL	AL069526 Drosophil
30	75	6.4	1085	12	CNS016YR	AL107373 Drosophil
c 31	75	6.4	1190	12	CNS02ON7	AL206908 Tetraodon
32	74.8	6.4	876	12	CNS009G1	AL053529 Drosophil
c 33	74.8	6.4	1203	12	CNS015WU	AL106008 Drosophil
34	74.6	6.4	307	12	CNS00A3W	AL054893 Drosophil
c 35	74.4	6.3	1027	12	CNS02T50	AL212733 Tetraodon
36	74.4	6.3	1101	12	CNS00FYG	AL071206 Drosophil
37	74.2	6.3	798	9	AI909641	AI909641 PM-BT217-
38	74	6.3	996	12	CNS00FUH	AL071063 Drosophil
39	73.8	6.3	928	12	CNS00DKY	AL071865 Drosophil
40	73.6	6.3	1101	12	CNS004ZW	AL055440 Drosophil
41	73.6	6.3	1101	12	CNS00LT2	AL078714 Drosophil
42	73.4	6.3	905	12	CNS00KHX	AL077798 Drosophil
43	73.2	6.2	987	12	CNS014PQ	AL104456 Drosophil
c 44	73	6.2	1101	12	CNS003BD	AL064091 Drosophil
45	72.8	6.2	1001	12	CNS0155H	AL105023 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

```
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             8.4%;  Score 99;  DB 12;  Length 1101;
  Best Local Similarity   20.2%;  Pred. No. 4.7e-08;
  Matches  143;  Conservative  301;  Mismatches  261;  Indels    4;  Gaps    2;

Qy      2 gtaggattggtggaatgggatcattcctaatcccttaatgacggtggcatgaacacaaag 61
          :: :: ::::|::::::   :  ::   |::     :|:::  ::|   :  | : :::
Db   1097 RWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWW 1038

Qy     62 caaagagaagttaggtcactcctcctttatatatatatatatatgcatgcatgaggacca 121
           :|  : ::|:::    | :  |  |:::||:: ::| ::: | : : : ||:|:||  :
Db   1037 WWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRR 978

Qy    122 tggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcactagttttt 181
          :|| :::: ::: :|::: : | :::| :  : ||: :::::::       :::: :::
Db    977 DGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDD 918

Qy    182 tgttggtgtgatttctcatgatgacgcgaaaattttatatatatatataatgaataatat 241
          :|  ::::::|  :: : :|: |:: : :: ::::::|: |: :  :|:|||:::::: :|:
Db    917 DGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWAD 858

Qy    242 gattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgacca 301
          ::: :: | : :  ::::: :  : ::|:|     |:: :  : | :: ||:: ::  :
Db    857 ADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKR 798

Qy    302 gttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaatttga 361
          :::  ||:::  | :  : ||  :  :|::::::|:|: ::|  : :: : ::: ::|
Db    797 ADDKRDAADDRDDAATWTTWTTTTTR--DTDDWKWKTDTWTRWAADRTWDRDDDDDRDRA 740

Qy    362 aaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggcta 421
            |   :::|: ::: :::| :::|: ::||::: :   ::  :| :    |:: :: ::
Db    739 GTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDR 680

Qy    422 atcaattattattaattaaaacgatagtatttttgataatttaattaaaatttttattgga 481
          || ::|:|::|: |::|  :   :::::::   :|| ::  | |  |::::|::|::  :
Db    679 ATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWD--W 622

Qy    482 tttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatata 541
          :::: |:|: : ::  : :| |  |::::   :|: |::: |: |  : :: |: : |
Db    621 AKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTT 562

Qy    542 atttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtat 601
          ||:|  ||: |::|:: : |    :|| : | :    ||| | |: | ||::: |  |||
Db    561 ATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTAT 502

Qy    602 aattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaata 661
          :| :    :| |   |: |  | || ||   : :    :  |  :|| |:    ::|||
Db    501 WAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTWWAATT 442

Qy    662 attattattttttctaataaagcacttcctaattgttaaaatatatgtc 710
           || :|:||::   || |  :    |    ||   :::: :|||||: |
Db    441 TTTTWTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATATKCC 393


RESULT   2
CNS00EVL
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR29B23"
                      /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN


  Query Match             8.0%;  Score 94;  DB 12;  Length 1101;
  Best Local Similarity   35.7%;  Pred. No. 3.6e-07;
  Matches  195;  Conservative  101;  Mismatches  247;  Indels    3;  Gaps    1;

Qy    192 atttctcatgatgacgcgaaaattttatatatatatataatgaataatatgattgattat 251
          ||:||||||: :| :: : ::||||:|:: | : | :|   :   : || |: ::: ||| |
Db    452 ATYTCTCAHTWTMMMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAAWAWWWWATTTT 511

Qy    252 tctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtgaccagttgtcaaat 311
          | : : |:|||:| | |::|: |:|||||  |    |   |      :    :  | |:|
Db    512 TWWWWWTWATTWTTTWAWWTWTAWTAAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAA 571

Qy    312 gaccactcgacttggggcatggtgatttttcaaatcacaactcaatttgaaaactaaaat 371
            :     |     :|    ||:|:   |||   ||   |  ||    : ||||||
Db    572 TTTAWAAWWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAAT 631

Qy    372 taaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaattatt 431
          |   ||  ||||| | :|||    || |  :   ::              | || ||:
Db    632 TTTTAATAATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWA 691

Qy    432 attaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaatgaa 491
          |  ||  ||||  |:| :| :  | |||: | |:||||||:  ||    | :   :  :|
Db    692 AAAAAAAAAAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWA 751

Qy    492 ctcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataatttagaaat 551
            ||::    | |:  :   ::|| :|::: |||:   || : |||:||: :| | |
Db    752 TAATATWWATATATAWTTWTAWWWATWWAWWWTATAW---ATAWAATAWAAWAWAWATAA 808

Qy    552 ataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgggcta 611
          |||:|| : |:|  : :|:|| :     |    : |||:|:|| |  |:||||     |:
Db    809 ATAWATAWATWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATW 868

Qy    612 gatttaacagctattattcaaactggccaggacaattattaaaattaataattattattt 671
            :|| |  ||: :|| : : |    |  | |   |: :| |: ||:||  |  |
Db    869 ATAAWAAAAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAA 928

Qy    672 tttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaaattt 731
            |  :|:::    ||  || |  : : ||::|| : | :: : : |||:| : ||:|
Db    929 AATAAWAWWWTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWT 988

Qy    732 atttgt 737
```

```
          || | |
Db    989 ATATAT 994


RESULT   3
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN


  Query Match             8.0%;  Score 93.4;  DB 12;  Length 1101;
  Best Local Similarity   33.7%;  Pred. No. 4.6e-07;
  Matches  216;  Conservative  120;  Mismatches  302;  Indels    3;  Gaps    1;

Qy     177 ttttttgttggtgtgatttctcatgatgacgcgaaaattttatatatatatataatgaat 236
           |::| | |:  | |  ||: |  :    |     ||:|  | ::|:|||||:||::  ||
Db    1100 TWWTATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTAT 1041

Qy     237 aatatgattgattattctctgtaattttgtgaa---atagattaaaacagctcaatgtga 293
           :||||  |   ||:|    | ||::||  | ||   ||: |:||::| |  | :|| : |
Db    1040 WATATATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAA 981

Qy     294 ggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaact 353
             |  : ::   :||| :  |           ||  : :::|:| |  |      |
Db     980 WTATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTT 921

Qy     354 caatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgg 413
           :|| : |  |:    || :: : |   : :: |:|: ||| ||   |  : | :
Db     920 TTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATA 861

Qy     414 gttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatt 473
           |    |  | : :||||::||  |:| :  ||: : :|||:  : ||:|| ||| : :|
Db     860 TTAWATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWT 801

Qy     474 ttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttat 533
           :| :|   :|  :   |: : :: |: :| : |: | |  |::  |   |::|:  |: |
Db     800 WTTWTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTT 741

Qy     534 gtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaa 593
           |  ||| :|||  |: |||: || |||:    |:   | |    ||| | |  | :|||
Db     740 WTTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAA 681

Qy     594 acacgtataattgggctagatttaacagctattattcaaactggccaggacaattattaa 653
           |  ||    :    |: :||    |   |:|| | ||| |    |  |   |  : ||
Db     680 TTATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAA 621

Qy     654 aattaataattattatttttctaataaagcacttcctaattgttaaaatatatgtctaa 713
           || || |  ||  ||||  |: :| | |   | |     |:: |: |||| : |   : |
Db     620 AAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTA 561

Qy     714 acactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaata 773
           : |   ||| | :||| ||| | :    ::   | :|  : : :          :::|:
Db     560 TTWWTTTAAATTATWTTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTW 501

Qy     774 aattagtgcataaaatataatggattggtggtctgtgaaaa 814
           ||     :  ||:| |: |  : :  |  : ::::|:|:
Db     500 TTTTTWATTWATAAWTTTWTTTWWAWATTWWKKKKKKAWAD 460


RESULT   4
CNS00EO7
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29P01"
                     /note="end : TET3"
BASE COUNT      366 a     66 c    104 g    351 t    214 others
ORIGIN


  Query Match             7.8%;  Score 91.4;  DB 12;  Length 1101;
```

```
  Best Local Similarity   40.0%;  Pred. No. 1e-06;
  Matches  183;  Conservative   70;  Mismatches  203;  Indels    2;  Gaps    1;

Qy     343 aaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattagg 402
           |||||| :  | |||     ::    |||| |:::| : |||   ::: || |  | :|
Db     523 AAATCAAWTWTAAATAATTWWTTATAAATAAWWWATKTTTTTATWWWATATAAAAAWAAT 582

Qy     403 ttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatt 462
           ||||||      ||     || :| : ::|  ||  :: |  |:: :    || |||:||
Db     583 TTAATTTTTWATTTTATATATTWAAWGWWAAAAAAAWWTATAAWHTWTAAATTTATAWTT 642

Qy     463 taattaaaatttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaat 522
           :|: |   |  | : |  |||| |:| |: |    | | |:   ||: |    |:: || |
Db     643 WAWATTTTAAATAWATTTATTTAAWTTAWWTWTTTAAWAWTTTAAWTATTAAARWWAATT 702

Qy     523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
           :|| |:   |: | : |||:  | ||| ||||||  ||    |||||:: :  |   |
Db     703 WAAAAWAAAAWTTTWAATAAWAAATAAAAATAAATTTTTTTTATTTAAWWAWTTAAAATT 762

Qy     583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggccagg 642
             |:||:|||:     : ||       : ||  :|    :| || |  |  :    |
Db     763 AWTWATWAAWAWTTWTAWAAAAAAAATNWTATAAWATTWWKAATAATWTATAWWTTTAWA 822

Qy     643 acaattattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaa 702
           : |:| ||  :||||: || ||:||| :| |||:|      |   |||:  |  ||
Db     823 TATWATRTAAATTWTAATWYTTTTTWTTTAWTTTAAWATTTTTTTAWWTAAWAWTGTAAT 882

Qy     703 tatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggtgagaggt 762
           :| ||   :  | |  ||:||| :  |:||  | :   | :||   | |      |  |
Db     883 WAAATTAWWTTAAA--AAWAATGWTTTWTAAATTWWGTTAWTTTTTATTTTWAAATTGTA 940

Qy     763 gctgacaaataaattagtgcataaaatataatggattg 800
             | | |||::| |  ||  |::||||:|::|| |||:
Db     941 TATAAAAAAWWATTATGTATAWWAAATRTWWTGTATTK 978


RESULT   5
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             7.6%;  Score 89.2;  DB 12;  Length 1101;
  Best Local Similarity   18.8%;  Pred. No. 2.5e-06;
  Matches  128;  Conservative  280;  Mismatches  272;  Indels    0;  Gaps    0;

Qy     361 aaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttggct 420
           ||||  |  :: |:|:|| |  |  :  | :||| :|:||  || |:| |
Db     422 AAAATAATTWWAAWAWAAAAAATTWWAAAAWAAAAWAWTAWWTTTAWTWAAAAAAAAAAA 481

Qy     421 aatcaattattattaattaaaacgatagtatttttgataatttaattaaaatttttattgg 480
           | |   || ||:|| |:| :|   :|  ::||    | ||    ||  :|||:   : |
Db     482 ATTWTTTTTTTWTTTAWTTWATAWWTTWWWTTAAAWAAAAAAAAAAAWAAAWAAAWATAA 541

Qy     481 atttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatat 540
           ||:| : :   :| :| :| |  :|:||:  : :  :  : :| :| |:|::: |  |:
Db     542 ATWTWWTWWTTYTTWAAWATAAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHA 601

Qy     541 aatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgta 600
           :|:||: :: : :|::: :|::: :: :::::| :  | ::::|| : :: ::::::|: :|
Db     602 WAHTTWYHWYHTYAMWHWMTWHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHA 661

Qy     601 taattgggctagatttaacagctattattcaaactggccaggacaattattaaaattaat 660
            |::|    :  |:::| :: ::  :   : :  :  ::   ::::||:: : ::: |:
Db     662 AAWWTTHTWWTHAYHWATYHYYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAH 721

Qy     661 aattattatttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaa 720
               : : ::: :||| | :  ::::::   :|| : :|:| : ::  ::||| |  |:
Db     722 YYMWYYMWAYYWMYCTACTYHYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAW 781

Qy     721 taataaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagt 780
              :: ::||:  :: : :::: :|  :: |   :   :  : |: : ::: : :|:: :
Db     782 ATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYH 841

Qy     781 gcataaaatataatggattggtggtctgtgaaaagacaggtggaggacaagccacctctc 840
             ::::  |::|::  :|:  :  : | : :| : :      :  :::: ::: |: ::
Db     842 HTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHH 901

Qy     841 tcaagtcaaaaggccatttcacaaccaacccaaatgggaacccaccaccgttccccgcca 900
           : ::  |::::   |:::: :::: |::: ::::::   :::  ::  :    ::|| :::
Db     902 HTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMCCMMHH 961

Qy     901 ttaaaatccctaatctcaccaaccccaactccacagattcttcaccaaacgcaactgattt 960
           : :: : ::::: |:: |  : :  :::::|  ::||:: :  |    :  : :||
Db     962 HCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTA 1021

Qy     961 ttcaatcaatgttttccctatactaccccccccaacaactccataatacccaatttgtcct 1020
           |: :: :::  : |:   :::: ||   | |::: : :|:| :  :: ::::: :    :
Db    1022 TWWWWMHWWAHWATWWWWWWWWATAWWACTCHHWTWYHTHCTWYYHHTYHMWWAWWMAWH 1081

Qy    1021 ttcaccaacccccgtcctcc 1040
           ::::: : :  :: :::|:|
Db    1082 WHHMYAHYHWAHHCWYYTMC 1101


RESULT   6
CNS00EO7/c
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
```

```
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29P01"
                     /note="end : TET3"
BASE COUNT      366 a     66 c    104 g    351 t    214 others
ORIGIN


  Query Match             7.2%;  Score 84.8;  DB 12;  Length 1101;
  Best Local Similarity   37.6%;  Pred. No. 1.5e-05;
  Matches  219;  Conservative   79;  Mismatches  283;  Indels    2;  Gaps    1;

Qy     150 ttgtgattgagatatgtccagcactagtttttttgttggtgtgatttctcatgatgacgcg 209
           || |  || : |:|:   :: :  |  | :| | ||  | | :  | |:  :       :
Db    1036 TTTTTTTTTWMAWAYACAMMAYTWTTTTAWTATTTTTTTTTTTMAATATYCWATATTTTMA 977

Qy     210 aaaattttatatatatatataatgaataatatgattgattattctctgtaattttgtgaa 269
           | |  ::    | ::|||| |    ::  | | ||  |  ||| :     |     | |
Db     976 ATACAWWAYATTTWWTATACATAATWWTTTTTTATATACAATTTWAAAATAAAAAWTAAC 917

Qy     270 atagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttggggc 329
           ::|  |||:|| :  | : | | |  : |    :  |  |:|    : |  |
Db     916 WWAATTTAWAAAWCATTWTTTTTAAWWTAATTTWATTACAWTWTTAWWTAAAAAAATWTT 857

Qy     330 atggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatta 389
           |   | | :    |::    :| |  |: | | |  ||||::|| |:|  ||| : | |:
Db     856 AAAWTAAAWAAAAARWATTAWAATTTAYATWATATWTAAAWWTATAWATTATTMWWAATW 797

Qy     390 ttaaattattaggttaattcacgggttggctaatcaatt--attattaattaaaacgata 447
           ||| |: |||   || :|  |   ||   :|:| | ||  |:|::| |   |||  | |
Db     796 TTATAWNATTTTTTTTWTAWAAWTWTTWATWAWTAATTTTAAWTWWTTAAATAAAAAAAA 737

Qy     448 gtatttttgataatttaattaaaattttattggatttgaatgaactcaattacatcacaa 507
           |  |||| ||  :||| |:|||:||||:||  :  | :::      |:||| |: :  |
Db     736 TTTATTTTTATTTWTTATTWAAAWTTTTWTTTTWAATTWWYTTTAATAWTTAAAWTWTTA 677

Qy     508 aaaacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaaccttt 567
           ||:|  |||:  || |||  :  | |   ||:|:|:| :|| |||| ||:|:  |  |  ||
Db     676 AAWAWWTAAWTTAAATAAATWTATTTAAAATWTWAAWTATAAATTTAWADWTTATAWWTT 617

Qy     568 taaatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctatta 627
           |   | :|   :      || || :|| |  || | |:    || ||:::| |   |
Db     616 TTTTTWWCWTTWAATATATAAAATWAAAAATTAAATTWTTTTTATATWWWATAAAAAMAT 557

Qy     628 ttcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcact 687
           ::   | |    |  |    | | :|: | | |  |  ||||||:|| |   ||:

Db     556 WWWTTATTTATAAWWAATTATTTAWAWTTGATTTTTATTTATTTWTTTTTTAAAWTTTTA 497

Qy     688 tcctaattgttaaaatatatgtctaaacactaataataaaatt 730
           |  || :  |||  |  : | : ||     |  | |||   ||
Db     496 TTATATWATTTATTAATWTTAWTTATYTTTTTTTTTATATCTTT 454


RESULT   7
CNS02ON7
LOCUS       CNS02ON7                1190 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            153P04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL206908
VERSION     AL206908.1  GI:7865727
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1190)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1190)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1190)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1190
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="153P04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG153DH02SP1~end :
                     PUC-Ori"
BASE COUNT      373 a    229 c    141 g    312 t    135 others
ORIGIN


  Query Match             7.2%;  Score 84.8;  DB 12;  Length 1190;
  Best Local Similarity   36.9%;  Pred. No. 1.5e-05;
  Matches  208;  Conservative   68;  Mismatches  288;  Indels    0;  Gaps    0;

Qy     170 gcactagttttttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatat 229
           :: :: :::::|::|||: |   | :  | |  |  |  |    | ||    | | | | | |
Db     439 SSSSBKKBBBBTKKTTKTTTAAAAWAATTAAAAATAAATAATAAAAANAAAAAAAAAAAA 498

Qy     230 aatgaataatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaat 289
           ||  || ||:|  :|| | :||| |   |||   | |  ||:| ||| ||| |  |   |
Db     499 AAAAAAAAAWAAAWTTAAAWATTTTTAATAANAATTTTNAAWATATTTAAAAATATTTTT 558

Qy     290 gtgaggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcac 349
           | |  |    | :| |    :           :|         : ||:::: :||  :
Db     559 TTAAAATTTWAATWTTTATTTWTTTWTANNTTTWTATATNTATTWWATWWWWAWAAAAWA 618

Qy     350 aactcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattc 409
```

```
          |:    :| | || | :|: |  ||||   |  :| :      ||| : ||  | |
Db    619 AWAAATWTATAAATATWAWTAAAAAAATTWAAAWAAWAATAATAATAWNTAWATAATAAW 678

Qy    410 acgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaa 469
          |    |    ||   :|| ::: | ||  |||    ||: ||  | |    || ||| |
Db    679 AAAAATATTATATAAWATAWWWAATATATAATTATATWATANATATATATATATAAAATA 738

Qy    470 aattttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatc 529
          ||| ||||   |||   | |  |  |||   : | :: | |  : || | | ||:|:
Db    739 TATTATATTTTTTTTTTTTNATATATATTTNTWAAWWWTATATAWTATAATANTAWTWAW 798

Qy    530 ttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttata 589
          | :| | :| | ||     |:::: | :  ||      |  ::| |    | | : |  ||
Db    799 TAWTNTTWTTTTATAATATAWWWWTATWTTTTTTWTATATWWTWTATTWTATAWATATTA 858

Qy    590 aaaaacacgtataattgggctagatttaacagctattattcaaactggccaggacaatta 649
          ||||      | |||    :| ||  :|       |||  ::          |  ||
Db    859 TAAAAATATNTTTATTAAAWWATATAAWAATATNTATATANWWTANAATATAWAWTATAT 918

Qy    650 ttaaaattaataattattatttttttctaataaagcacttcctaattgttaaaatatatgt 709
          || |::  |||  ||| : | | | :|||| | ||: |  |:  | |  |||:|||
Db    919 ATATAMWATATATATATAWATATATAWAATATAWCAYATNATWTATATATAAAWATANAA 978

Qy    710 ctaaacactaatataaaatttat 733
            |    ||::|| |||   |||
Db    979 TATATAYATAMWAACAAATAATAT 1002
```

```
RESULT   8
CNS0021J
LOCUS       CNS0021J                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061936
VERSION     AL061936.1  GI:4940214
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR05N11"
                     /note="end : TET3"
BASE COUNT      631 a      7 c     28 g    289 t    146 others
ORIGIN
```

```
  Query Match             7.2%;  Score 84;  DB 12;  Length 1101;
  Best Local Similarity   45.3%;  Pred. No. 2.1e-05;
  Matches  184;  Conservative   35;  Mismatches  184;  Indels    3;  Gaps    1;

Qy    343 aaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattagg 402
          |||  | ||   ||    ||||  ||||  |||||| |   | |  |  ||  || |
Db    464 AAANAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAANAAAAATATAATTTA 523

Qy    403 ttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatt 462
          :|  || :    ||   |  |   || || ||  |:  ||   || | |||   |:|| |
Db    524 WTTTTTTWTTAATTAWTTTTTTTTTTTTTTTTTTTWTTAATTTTAATTTTTAAWAWAAAT 583

Qy    463 taattaaaatttttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaat 522
          | | |||||: :|:||  :||| |||  | : ||  | |      |||:  |: |    |
Db    584 TTAATAAAAWAWTWTTTAWTTTTAATWTAAWWAAAAAAAAWTTTTAAAWAWTWTTTTTTT 643

Qy    523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
          |: ||  | || : | | |: | : |: |:|  || : |:|  |: || | |  |  :|
Db    644 TWTTAATTAATAWAAAAAAWATAWAAWTTWTTTATWWATWAAATWAAATTTTAAAAAWTT 703

Qy    583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggcca-- 640
          | |||:       :  || ||:|    ||  ||||| |  |: : :: ||| |    |
Db    704 TTTTAWTTTTTTTWTTTAAAAWTTTATTAWTTTTAAAAWATWAWTWWTAAATTTTTTAWA 763

Qy    641 -ggacaattattaaaattaataattattatttttttctaataaagcacttcctaattgtta 699
           |: |  :||||||  | | | | |||||||| |||| ||  |      ||     :
Db    764 TAWTAAWATTAWAAAATTTTTTAATTTAATTTTTTTTAATTAAAAAAAAAAAAAAAAAAW 823

Qy    700 aaatatatgtctaaacactaataataaaatttatttgtgtatcttt 745
          |||||  |    |||| |  || |: ||::||| |||   |   ||:
Db    824 AAATAAAAAAATAAAAAAAAAAAWAAAWWTTTTTTTAAATTATTTW 869
```

```
RESULT   9
CNS003BD
LOCUS       CNS003BD                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064091
VERSION     AL064091.1  GI:4941847
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
```

```
                     /clone_lib="RPCI-98"
                     /clone="BACR08K08"
                     /note="end : TET3"
BASE COUNT      395 a    120 c    103 g    334 t    149 others
ORIGIN
```

```
  Query Match             7.0%;  Score 82.2;  DB 12;  Length 1101;
  Best Local Similarity   39.2%;  Pred. No. 4.4e-05;
  Matches  157;  Conservative   57;  Mismatches  186;  Indels    0;  Gaps    0;

Qy     343 aaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattagg 402
           |:|| | ||   ||    ||||  || || |||||| | ||| |  :|: || |||
Db     675 AWATTAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAATTATAAAWTWTAAATATATAW 734

Qy     403 ttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatt 462
           |:  |  |     |     ||  |: ||:| ||| ||| |   | | |||: |:  | ||
Db     735 TWTTTAWAAATATATTAAAAAAAWATAWTTTTATTTATATTAAAAATATWATWATTTATA 794

Qy     463 taattaaaatttttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaat 522
           ||  : :|  :|| :|  |  |  :  |: | |||: :|| | |:|     :||| |||
Db     795 TATNWNWATAWTTTWTTTAAATTTWATAWTTAAATWTMATTAAAWATTNTAWAATAAAAA 854

Qy     523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
           :|| |   ||    | |||||:|| : :   :: | :  |   :| ||:|    |
Db     855 WAAAAAAATAATAAAAATAATWTATWTWATAWWTTTWAAATAAWTAAAWTAAWAAAAAAA 914

Qy     583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggccagg 642
             ||| :| ||| |   |: |:|    |:  | :|| :    | : : :||
Db     915 AATTAAWATAAAAATWAAWTAWTTWTTTMTYTAWAAAWYWATTAWAWAWAATATTTTTAT 974

Qy     643 acaattattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaa 702
           | :||:|||| ||| :: ||  |  | |     :||   :  | : :  |: :  :||||
Db     975 ATWATWATTATAATAWWAAAAAAAAAATAAAATWAAATNWAWAWWAYAAAWAYACWAAAA 1034

Qy     703 tatatgtctaaacactaataataaaatttatttgtgtatc 742
            :|:| | :| | :  || |: |: |  |  |:   :|::
Db    1035 AWTMTATWYATAWMAAAAAAMAAHTANATTATWAAAWAWM 1074
```

```
RESULT  10
CNS01G8P
LOCUS       CNS01G8P                 910 bp    DNA     linear   GSS 01-JUN-2001
DEFINITION  Anopheles gambiae GSS SP6 end of clone 06E14 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL142826
VERSION     AL142826.1  GI:7000944
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Anopheles.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 910)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="06E14"
                     /clone_lib="NotreDame1"
                     /note="end : SP6"
BASE COUNT      376 a     83 c     91 g    294 t     66 others
ORIGIN
```

```
  Query Match             6.9%;  Score 81.2;  DB 12;  Length 910;
  Best Local Similarity   42.8%;  Pred. No. 6.8e-05;
  Matches  188;  Conservative   30;  Mismatches  221;  Indels    0;  Gaps    0;

Qy     349 caactcaatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaatt 408
           |:  |  | || || | : |    ||||||  |   | :|| |:|:  || :|  ||:|
Db     472 CWTTTTTAATTAAATAAWTATTNNAAAAAAWAAAAAWWTTTTWAWTATAAWATTTTWAAA 531

Qy     409 cacgggttggctaatcaattattattaattaaaacgatagtatttttgataatttaatta 468
           |     |   |  | || :| |:|||||  |||  | | |   | |  :|||| ||  :
Db     532 AAATATATATWTTTTTAAAWAATWTTAATATAAAAAAAATTTAATATWTWAATTAAAAAW 591

Qy     469 aaattttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatat 528
           :|| ||| ||   | | |:| |  : |   |      ||||||   ||  |||    |||
Db     592 WAAATTTTTTATTATTAWTAATTWAANAAAATAATAAAAAAAATAAAAAAAAAANTTAT 651

Qy     529 cttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttat 588
           :: |    ||||   :| |||  |  || || |:  | ||||       |:  | |  |
Db     652 WWWTTTAATTATATAAWAAAAAATTTTATAATAAWTTTATAAAAAAWATTTWAATWTACT 711

Qy     589 aaaaaacacgtataattgggctagatttaacagctattattcaaactggccaggacaatt 648
           | |:| |   | || |     | | : || |  || || | |||      | | |||
Db     712 TATAWATAAAAAAAAAATATTWAAAAAWAAAAATATAATAATAAAATAANAAAAAAAAATA 771

Qy     649 attaaaattaataattattatttttttctaataaagcacttcctaattgttaaaatatatg 708
           |   :||| ||   |:|:      || |:|| |  |  : ||    | |   :
Db     772 NAAATTNWTAAAAAAATTWAWNANAATTAYTWAACAAWNTNTWTATAAAAATATNYWTAA 831

Qy     709 tctaaacactaataataaaatttatttgtgtatctttggcagtaggtgagaggtgctgac 768
           |||:  ||||: | |  || || |   ||| |||   | |       |  |  | |
Db     832 AATAAWWTTTAATWTTNATTTTAATATCATTATTTTTATAATTTATATTAAAATAATTAT 891

Qy     769 aaataaattagtgcataaa 787
           |  ||||| :   |||||
Db     892 ATCTAAATAWTATAATAAA 910
```

```
RESULT  11
CNS010MP/c
LOCUS       CNS010MP                 734 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
```

```
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04L20"
                     /note="end : T7"
BASE COUNT      288 a     62 c      2 g    211 t    171 others
ORIGIN
```

```
  Query Match             6.9%;  Score 80.6;  DB 12;  Length 734;
  Best Local Similarity   33.3%;  Pred. No. 8.9e-05;
  Matches  223;  Conservative  119;  Mismatches  324;  Indels    4;  Gaps    1;

Qy      54 acacaaagcaaagagaagttaggtcactcctcctttatatatatatatatgcatgcat 113
           ::|:::: ::|| | || :||  :: : |   :::| | | | :::| :| :  | ::||
Db     725 VVAMVMMMHMAAAAAAAAWTAAMHSGSSCKGGMHHTTTTTTTTWWWAAWTGKTGAVKMAT 666

Qy     114 gaggaccatggctatgatgaaggttaatagaggtagttgtgattgagatatgtccagcac 173
           |  :  |  | ::||       | |  | :  |   | |:| ||  | | |   || |
Db     665 TAATWTAAMTGGYMTGKAMTTWMTCATAAMMAATGAATTTKAATGTWAAAAGAAAAGAAA 606

Qy     174 tagttttttgttggtgtgatttctcatgatgacgcgaaaatttatatatatatataatg 233
           |   | |   | |||  ::|: |  :| :   :   |||  |:: :   |   :
Db     605 AAWMGTATGMATAGTGATRWTWATGTKGTDTTTSATTAAAAATWDGDGTAARTGWGGTAT 546

Qy     234 aataatatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtga 293
           | : | ||:  ||  :|||   :| :| |::||:    | |  | |     | : |  |:
Db     545 ATWTAAATKGATGGAKATTWAAKGGWAGTKKGTKTTTGATAAAATAKTTTTTTDTTTGGR 486

Qy     294 ggtgaccagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaact 353
           ||  |   :|      |      |    |  ||:|        |||    | || |  | :
Db     485 GGGAATTTKTATRNTATAATTTAAAAAAAATGDGTGTGAAGTATTAAAAATATAAAGAAW 426

Qy     354 caatttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgg 413
           |||   || | |    :|    | | | |      ::||  :  :| |    |:|
Db     425 TAATAAWAATATTTGTGWTTTTTATGTTGTTATAAWWTATKWATWTTGTAATAKTWTTTT 366

Qy     414 gttggctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatt 473
           |  : |: |  :|  |||:|::||: |   :|  | |::||:|||:|||  |   : ||
Db     365 ATGTKTTWTTTTWTATTTAWTWWTTWTATTTWTTTTTTWWTTKATAWTTTTGTATTWTTT 306

Qy     474 ttattggatttgaatgaactcaattacatcacaaaaaacct----aatcaaattaatatc 529
           |:  :  |  | |||  : :   |:|  |       ::  |     || :::::: | |
Db     305 TDTGKWWAAATTAATATWGWTTTTKATTTTTTGTTGWWTTTTWTKTATAWWWWWWKTTTT 246

Qy     530 ttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttctcttata 589
           |::: |  | |: ||:  |||:  | || ||: :  :||:  | :   ||: |  :| |
Db     245 TKWWATATTKTWTTTKTTAAAWTAATATAATWTWTTWTTWTTTTWAWKWTTWATWAWAAA 186

Qy     590 aaaaacacgtataattgggctagatttaacagctattattcaaactggccaggacaatta 649
           || :  | ::::  |      |  |||||  :  |  | |  :  ::   |: | |||:|
Db     185 AATWTRAKKWWWTTTAAAAAAATTTTTAAATKAAAAAAATATWTTKDAAAAKAAAAATWA 126

Qy     650 ttaaaattaataattattatttttttctaataaagcacttcctaattgttaaaatatatgt 709
           ||  :|:::||::| |::|||||| :||    | | :|  |  :| |:|  :|||||
Db     125 ATATTWTKDWTAWWTTTWWTTTTTTKKAAATTTGAAWKTKTTTTWTATWATTWTATATAA 66

Qy     710 ctaaacacta 719
           || | | ||
Db      65 KTATATAATA 56
```

```
RESULT  12
CNS020K7
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1-end : T7"
BASE COUNT      383 a    169 c    165 g    262 t    113 others
ORIGIN
```

```
  Query Match             6.8%;  Score 79.6;  DB 12;  Length 1092;
  Best Local Similarity   42.1%;  Pred. No. 0.00013;
  Matches  164;  Conservative   54;  Mismatches  160;  Indels   12;  Gaps    1;

Qy     343 aaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattagg 402
           |||  | || : :| |: ||::  |||| |:||:: |:||:  || |:|:::::::|
Db     713 AAAAAAAAAAWTWAATWAAAWWAAAAAAWAAWAAWWWAWTTWTTTTTTWAWWWWWWWAAA 772

Qy     403 ttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatt 462
           :|| : :    :   ||:: | : ::| :|| || |||  | |  :||| :   :
Db     773 AWAAAWWWTWTTAWTTTTAWWWATWAWWTTWTATTTTAAAWAAAANAWTTTAWTTAWTAA 832

Qy     463 taattaaaatttttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaat 522
           || ::|||| | :|  |    |:|  |   | |:: ||   ||||||   | | |||
Db     833 AAAAWWAAATAATTWTTTAAAAWAWTTTAWAAATTWWAATWTAAAAAAAAAAAATTAAAAA 892

Qy     523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
           :| :| :| | |  |:|:|::|: ||||:|            |||   | |    :||
Db     893 AWAAWTWWTTTTTTTTWTWAWWTWTAAATWT-----------TTTTTTTTTTTWWTTT 940

Qy     583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggccagg 642
           | :|| :||||| |   | |  |     | | ::   |  |||| || |||      |
Db     941 TTWTAAWAAAAAAAAAAAAATATAAAWNAAAAWWTTAATTTATTTTTAAAAAAAAAAAAA 1000

Qy     643 acaattattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaa 702
           |  |  ||  |||: |||||    | | | |  ||| |||  |     ||| |   |||:
Db    1001 AATAAAATATAAAWAAATAAAATATTTATATATTAAAAAAAAAAAAAAATAAATTAAAAAW 1060

Qy     703 tatatgtctaaacactaataataaaattta 732
           || |     ||| |  || || |||| |||
Db    1061 TAAAATAAAAAAAAAAAAAAAAAAAATTA 1090
```

```
RESULT  13
CNS003BB
LOCUS       CNS003BB               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR08G08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064089
VERSION     AL064089.1  GI:4941845
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08G08"
                     /note="end : T7"
BASE COUNT      430 a    118 c    167 g    261 t    125 others
ORIGIN


  Query Match             6.7%;  Score 79;  DB 12;  Length 1101;
  Best Local Similarity   46.3%;  Pred. No. 0.00016;
  Matches  179;  Conservative   26;  Mismatches  180;  Indels    2;  Gaps    1;

Qy     366 taaaattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatca 425
           || | |:||| ||  |   :|||   |||  || |  : |||       ||   |:|  |
Db     650 TATAKTWAAATAAATTAATDATTTAAAAAAAATAATTWAAATAATTAAATTAAATWAAAA 709

Qy     426 attattattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttg 485
           |  | |||| || ||||  ||   |||||  |     :|||    | :  |    |: |
Db     710 AAAAATATTTATAAAAATTATTNAATTTTAAAATTAAWAATATTTAAWAAAAATAAWATA 769

Qy     486 aatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataatttt 545
           ||    | ||    :  | ||||||  |||| |||  |||:| :    | ||| | |||
Db     770 TATTTTTTAAAAATTWAAAAAAAAAAAAATAATAAAA--AATWTAWNTATTNATAAANTTT 827

Qy     546 agaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataatt 605
           || | ||:| ||:    |  ||||      ||| | | | |||||| |  || ||:|
Db     828 TTTAAAAAAAWTTATWTTNTTAAAAATTAATTATTTTTTTAAAAAAAAAAAAATAAAAWT 887

Qy     606 gggctagatttaacagctattattcaaactggccaggacaattattaaaattaataatta 665
           |    ||:| |  |||   : | : |    |    |  |   ||||  |: ||| |
Db     888 AATWTTWTATTWATATWTATNTNWTATWTTAAATATATAATGTTAAAAAAAAAWAAATAA 947

Qy     666 ttatttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataata 725
           || :|||| |||  :|  |  :  : ||| :|: ::|||||    ||| |  || ||:|
Db     948 ATAAWTTTTTTAAANWAAWATAWAAWTATTTWTWTWWTATATAAAAAAAAAAAAAAAWA 1007

Qy     726 aaatttatttgtgtatctttggcagta 752
           |  || | |    || | |   | ||
Db    1008 TATATTTTATAGWAATATATAAAAATA 1034


RESULT  14
CNS04D0K
LOCUS       CNS04D0K               945 bp    DNA     linear   GSS 21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1-end : T7"
BASE COUNT      386 a    112 c     96 g    231 t    120 others
ORIGIN


  Query Match             6.7%;  Score 78.4;  DB 12;  Length 945;
  Best Local Similarity   41.5%;  Pred. No. 0.00021;
  Matches  192;  Conservative   49;  Mismatches  215;  Indels    7;  Gaps    1;

Qy     336 atttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaat 395
           |||:|: |||: : ||   ||  | |:|: :    | |||||| ||    |:  |: :
Db     490 ATTWTWAAAAWAWAAATAAAAAATAAWAWWWTTTTTAAAAAAAWATAATAAWATTWTWTA 549

Qy     396 tattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtattttt 455
           :|::|  ||:||        |    :|  |||    |  | || |||  : | |||||||
Db     550 WAWWATTTTWATATTTTTTATTTAAWAATAATATAAAAAATTTTAAAAAWAAATATTTTT 609

Qy     456 gataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaaaaccta 515
           |:|||  |: ||||| |  ||    ||  |:  :|   |: :: :: | :||:::  :|
Db     610 AAWAATAAAWATAAAAATAAATATATTTATAWAAWAWATAWAWWWWWTAWWAAWWWTAWA 669

Qy     516 atcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctc 575
           ||  ||  |||| :|:  | | |  | ||| |:|||: |  ::  | |:|
Db     670 ATATAAAAAATAAAAWAWAATAAAAATAAAATAAAAAWAAAWWAAAWWTWATAAWANAAA 729
```

```
Qy     576 gagtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaact 635
Db     730 AAAAAATATATAATAWAAA-------ATATAAATTAAATATAAAATAWATWWATAAATAA 782

Qy     636 ggccaggacaattattaaaattaataattattatttttttctaataaagcacttcctaatt 695
Db     783 AAAAATTTWTAATAATAAAAAWATTAAAWATTATWTATTANWATAAAANTTATAAWNTAW 842

Qy     696 gttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggt 755
Db     843 NTAATAAATAAAATWTAAAAAAAAAAAAWAAAATTATAAAAAATAAAATTATAAAAATAA 902

Qy     756 gagaggtgctgacaaataaattagtgcataaaatataatggat 798
Db     903 TATAAAAAAAAAAAAAAAAATWTTAAAAAAAATAAAAAAATAT 945


RESULT 15
CNS0039L/c
LOCUS       CNS0039L                 1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR08I10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063926
VERSION     AL063926.1  GI:4941783
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08I10"
                     /note="end : T7"
BASE COUNT      139 a     79 c    186 g    360 t    337 others
ORIGIN


  Query Match             6.7%;  Score 78.4;  DB 12;  Length 1101;
  Best Local Similarity   27.1%;  Pred. No. 0.00021;
  Matches  199;  Conservative  185;  Mismatches  343;  Indels    7;  Gaps    1;

Qy     343 aaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaattattagg 402
Db    1067 AARRAAAAWRTAWAWAWAATAAAKWKWRWAWRAAWWDWGATGDTWTADWGDADDRRAADR 1008

Qy     403 ttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttgataatt 462
Db    1007 RADRRRAADATARKRKKTGWTGARATRTKTRTRWTTWRAWAAGTRATKWGDTDARTWTGT 948

Qy     463 taattaaaatttttattggatttgaatgaactcaattacatcacaaaaaacctaatcaaat 522
Db     947 TTARTATGKKGTATWKGTTGKTTKKTTKRGGTKRKTKKAKAWAAAAAAAAAAAAAAWAWW 888

Qy     523 taatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcgagtttc 582
Db     887 TTATATAATDTKWRATAGWAWATTKADATADWAWADATKRTKTATRAATTDKDAAAAAAA 828

Qy     583 tcttataaaaaacacgtataattgggctagatttaacagctattattcaaactggccagg 642
Db     827 AARATGAAAARARAAAAAARAKWTDDRWAAARWDAAARDARGTGGTAARDGWGGKKAAAK 768

Qy     643 acaattattaaaattaataattattatttttttctaataaagcacttcctaattgttaaaa 702
Db     767 WTGRTTAAAADAWAAAAAAAAATWAWWWAATAAKTAAAAAATARWGWAAAATKWTKAAAAA 708

Qy     703 tatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggtgagaggt 762
Db     707 ADAWAAARAGDKTAAAAAAAAAARARAGRGGKRTRKKRARAAATAAAATAAAAAAAAAAD 648

Qy     763 gctgacaaataaattagtgcata-------aaatataatggattggtggtctgtgaaaag 815
Db     647 AWTAKKAAWAARTTTRTGGKRTGKTKGTGGKGAAWWAGTATKRKAAAAAKTWKAAAAAAA 588

Qy     816 acaggtggaggacaagccacctctctcaagtcaaaaggccatttcacaaccaacccaaat 875
Db     587 AAAGKRTGGKKKKKTRTBTRDWNBNAHAAATMWNAATBMTGTGKSWMMMATMDMTMTMKG 528

Qy     876 gggaacccaccaccgttccccgccattaaaatccctaatctcaccaacccaactccacag 935
Db     527 KRGGWRGGRKTKKVTTMMVRTGCGCCCCCMAKAKTTSGCCBTBTYCMMMAAACNCCMCMA 468

Qy     936 attcttcaccaaacgcaactgatttttcaatcaatgttttccctatactacccccccaac 995
Db     467 MTMAMMMCCAMMMMMCMCMAMACNANCTMACAAMCCAMCCACACMCCCMMACACMCMACA 408

Qy     996 aactccataatacccaatttgtcctttcaccaacccccgtcctccgtgccagccaattct 1055
Db     407 CMCCAACCCACATMCMACMCCCMMMTTMAACCAAATMAMAAAACCMTMMAACTCAAMMCC 348

Qy    1056 atatcagcaggaat 1069
Db     347 CTCTMAMMMAMACT 334


Search completed: July 30, 2002, 10:38:02
Job time: 5471 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 10:04:36 ; Search time 50.16 Seconds
(without alignments)
5739.284 Million cell updates/sec

Title: US-09-530-663B-5
Perfect score: 1172
Sequence: 1 tgtaggattggtggaatggg.....agcccgcaatggacgcgaca 1172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	61.8	5.3	636	4	US-08-998-416-1137	Sequence 1137, Ap
	2	61.8	5.3	665	2	US-08-883-795A-36	Sequence 36, Appl
	3	61.4	5.2	660	1	US-07-991-867B-32	Sequence 32, Appl
	4	61.4	5.2	660	1	US-08-107-755A-32	Sequence 32, Appl
	5	61.4	5.2	660	2	US-08-544-332-32	Sequence 32, Appl
	6	61.4	5.2	1511	1	US-07-991-867B-8	Sequence 8, Appli
	7	61.4	5.2	1511	1	US-08-107-755A-8	Sequence 8, Appli
	8	61.4	5.2	1511	2	US-08-544-332-8	Sequence 8, Appli
	9	61.4	5.2	4810	3	US-08-852-629-11	Sequence 11, Appl
	10	61.4	5.2	4838	3	US-08-852-629-15	Sequence 15, Appl
	11	61.2	5.2	837	4	US-08-998-416-288	Sequence 288, App
c	12	60	5.1	1431	4	US-09-316-083-2	Sequence 2, Appli
	13	60	5.1	6243	2	US-09-056-075-1	Sequence 1, Appli
c	14	58.8	5.0	636	4	US-08-998-416-1137	Sequence 1137, Ap
c	15	58.8	5.0	837	4	US-08-998-416-288	Sequence 288, App
c	16	58.2	5.0	615	4	US-08-998-416-186	Sequence 186, App
	17	57.6	4.9	2317	3	US-08-749-522-5	Sequence 5, Appli
	18	57.6	4.9	3974	4	US-08-467-504-3	Sequence 3, Appli
c	19	57.2	4.9	665	2	US-08-883-795A-36	Sequence 36, Appl
c	20	57.2	4.9	3095	6	5231168-1	Patent No. 5231168
	21	56.8	4.8	3095	6	5231168-1	Patent No. 5231168
	22	56.6	4.8	615	4	US-08-998-416-186	Sequence 186, App
c	23	55.6	4.7	2317	3	US-08-749-522-5	Sequence 5, Appli
c	24	55.6	4.7	3974	4	US-08-467-504-3	Sequence 3, Appli
	25	55.4	4.7	724	4	US-08-998-416-683	Sequence 683, App
c	26	55.4	4.7	854	4	US-08-998-416-534	Sequence 534, App
c	27	55.4	4.7	860	4	US-08-998-416-287	Sequence 287, App
c	28	54.6	4.7	663	4	US-08-998-416-191	Sequence 191, App
c	29	54.6	4.7	1186	2	US-08-731-722-5	Sequence 5, Appli
c	30	54	4.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
	31	53.6	4.6	854	4	US-08-998-416-534	Sequence 534, App
	32	53.6	4.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
	33	53.4	4.6	3926	2	US-08-731-722-1	Sequence 1, Appli
c	34	53.4	4.6	3926	2	US-08-731-722-1	Sequence 1, Appli
	35	53.4	4.6	3926	2	US-08-731-722-2	Sequence 2, Appli
c	36	53.4	4.6	3926	2	US-08-731-722-2	Sequence 2, Appli
c	37	53	4.5	658	4	US-08-998-416-595	Sequence 595, App
	38	52.8	4.5	658	4	US-08-998-416-595	Sequence 595, App
	39	52.4	4.5	2251	4	US-08-991-677-11	Sequence 11, Appl
c	40	52.4	4.5	6124	4	US-08-213-419B-3	Sequence 3, Appli
	41	52.2	4.5	701	4	US-08-998-416-701	Sequence 701, App
c	42	52.2	4.5	1422	1	US-08-319-704-5	Sequence 5, Appli
c	43	52	4.4	724	4	US-08-998-416-683	Sequence 683, App
	44	52	4.4	732	4	US-08-998-416-1036	Sequence 1036, Ap
c	45	52	4.4	732	4	US-08-998-416-1036	Sequence 1036, Ap

ALIGNMENTS

RESULT 1
US-08-998-416-1137
; Sequence 1137, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692RP

US-08-998-416-1137

Query Match 5.3%; Score 61.8; DB 4; Length 636;
Best Local Similarity 48.3%; Pred. No. 3.5e-05;
Matches 208; Conservative 0; Mismatches 217; Indels 6; Gaps 1;

```
Qy    370 attaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatta 429
          ||  ||  | |   | | |||| |||    |   || ||       |   | |||     |
Db    177 ATATAATTATAAAAATAATATTTAATATGAATACTATTTAGTCTATGTTCAAATTTTAAA 236

Qy    430 ttattaattaaaacgatagtatttttgataatttaattaaa------attttattggatt 483
          ||| | |||||||   || ||  | | || ||||  || ||      |||  || | |
Db    237 TTAGTTATTAAAATATTATTAGATATTATTATTTTCTTTAATAAATTATTAAATAGATTA 296

Qy    484 tgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtgatataat 543
          | ||| |     ||   ||     ||    | || ||| |||||| || | | ||   |
Db    297 TCAATAATTAATATATTATTTATTAATTGTTTATTAAAATAATATATTTTATTATTATAA 356

Qy    544 ttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataa 603
            |   |  | | | |   |   | || | |    ||| |  ||  |   |    |||||
Db    357 AGATTTAATTTATTTAAATATTGTAAATTATTATTTTTATTATAATATCTATTTTTATAA 416

Qy    604 ttgggctagatttaacagctattattcaaactggccaggacaattattaaaattaataat 663
            |   |     |     ||||      |    || |  |||||||||||||||  |
Db    417 ATATTATGTTGATTTATATTATTTAATCTTTTTATAAGAATTATTATTAAAATTAATTTT 476

Qy    664 tattatttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataa 723
            | | |  ||||| || |   | ||    ||| || ||  |  | | | |  |   || | |
Db    477 AACTTTAATTTCTTATTATTAATTTTTATATTATTTAATAAATTATATTCATTTTATTTA 536

Qy    724 taaaatttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgca 783
          |  | ||  ||  |  |  |        ||  | | |  |  | |  | | ||||| |  |
Db    537 TTTATTTATTTAATTAAATTAATTATTTAATTAATATTTTATCATTATTTAATTAATTAA 596

Qy    784 taaaatataat 794
          |||||||| ||
Db    597 TAAAATATTAT 607
```

RESULT 2
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398

; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-883-795A-36

Query Match 5.3%; Score 61.8; DB 2; Length 665;
Best Local Similarity 48.8%; Pred. No. 3.5e-05;
Matches 224; Conservative 0; Mismatches 232; Indels 3; Gaps 2;

```
Qy    336 atttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaat 395
          |||||  || | | |  |  ||    ||| ||    |||  ||   ||  || ||||||
Db     36 ATTTTATAATTAAAATATTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAA 95

Qy    396 tattaggttaattcacgggttggctaatcaattattattaattaaaacgatagtattttt 455
          ||||    |||||  |    ||    | | || ||||  ||||||||    |  || ||
Db     96 TATT--TATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAA 153

Qy    456 gataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaaaaacctа 515
          |||  ||  | | |||| || |   |||  |||       |||||  ||     | ||  ||
Db    154 AATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAAT-TA 212

Qy    516 atcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctc 575
          |   | |||  ||   || |  |||||||  | | || || |     |  |||| |  |
Db    213 AAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTA 272

Qy    576 gagtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaact 635
            | | | | | || ||| |       | |          | |  |  |||| |  ||
Db    273 AAATGTTTATAATTAAATATTTTATAATTAAAATGTTTATAATTACATATTTTATAATTA 332

Qy    636 ggccaggacaattattaaaattaataattattatttttttctaataaagcacttcctaatt 695
                  | |   | | || |||||||  || |||   | || |    ||   | |
Db    333 AAATGTTTATAATTACATATTTTATAATTAAAATGTTTATAATTACATATTTTATAATTA 392

Qy    696 gttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggt 755
                |||| |  | |   ||  | |||||| | | |   ||  | | |   | |
Db    393 AAATGTTTATAATTACATATTTTATAATTAAAATGTTTATAATTACATATTTTATAATTA 452

Qy    756 gagaggtgctgacaaataaattagtgcataaaatataat 794
            | |  |  | |   ||  || | | ||||   ||||||
Db    453 CATATTTTATAAAGTATTTATAATTACATATTTTATAAT 491
```

RESULT 3
US-07-991-867B-32
; Sequence 32, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-991-867B-32


  Query Match             5.2%;  Score 61.4;  DB 1;  Length 660;
  Best Local Similarity   48.0%;  Pred. No. 4.2e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  |||| | |||  |   | |  |   |    || |    |
Db     91 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 150

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db    151 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 210

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
              |||| |     | ||    | | |||   ||| | ||| | |||||  |||
Db    211 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 270

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db    271 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 330

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |   || ||| |   || | |  || |       | ||| |||| |||| ||   |
Db    331 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 390

Qy    668 atttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db    391 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 450

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
           |  || ||       |    |  |      |    | | | | ||||||  |  ||  |
Db    451 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 510

Qy    788 atataat 794
          | | |||
Db    511 AAAAAAT 517


RESULT   4
US-08-107-755A-32
; Sequence 32, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-107-755A-32


  Query Match             5.2%;  Score 61.4;  DB 1;  Length 660;
  Best Local Similarity   48.0%;  Pred. No. 4.2e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  |||| | |||  |   | |  |   |    || |    |
Db     91 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 150

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db    151 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 210

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
              |||| |     | ||    | | |||   ||| | ||| | |||||  |||
Db    211 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 270

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db    271 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 330

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |   || ||| |   || | |  || |       | ||| |||| |||| ||   |
Db    331 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 390

Qy    668 atttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db    391 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 450

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
           |  || ||       |    |  |      |    | | | | ||||||  |  ||  |
```

```
Db    451 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 510

Qy    788 atataat 794
          | | |||
Db    511 AAAAAAT 517


RESULT   5
US-08-544-332-32
; Sequence 32, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-544-332-32


  Query Match             5.2%;  Score 61.4;  DB 2;  Length 660;
  Best Local Similarity   48.0%;  Pred. No. 4.2e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  |||| | ||| |   | |  |   |    || |    |
Db     91 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 150

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db    151 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 210

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
          | ||||  |     | ||    | | |||   ||| | ||| | ||||||  |||
Db    211 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 270

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   | | |  | | |   | | |    | |||||
Db    271 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 330

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || | |         | ||| |||| |||| ||   |
Db    331 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 390

Qy    668 atttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db    391 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 450

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
          |  || ||       |    | |       |    | | | | |||||| |  ||  |
Db    451 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 510

Qy    788 atataat 794
          | | |||
Db    511 AAAAAAT 517


RESULT   6
US-07-991-867B-8
; Sequence 8, Application US/07991867B
; Patent No. 5476781
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5476781el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-07-991-867B-8


  Query Match             5.2%;  Score 61.4;  DB 1;  Length 1511;
  Best Local Similarity   48.0%;  Pred. No. 5e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  ||||  | |||  |   | |  |   |    || |    |
Db    942 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 1001

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db   1002 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 1061

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
             | ||||  |     | ||    | | |||   ||| | ||| | ||||||  |||
Db   1062 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 1121

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db   1122 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 1181

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || |         | ||| |||| |||| ||   |
Db   1182 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 1241

Qy    668 atttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | |||||  ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db   1242 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 1301

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
            |  || ||       |   |  |     |     | | | | |||||| |  ||  |
Db   1302 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 1361

Qy    788 atataat 794
          | | |||
Db   1362 AAAAAAT 1368


RESULT   7
US-08-107-755A-8
; Sequence 8, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-107-755A-8


  Query Match             5.2%;  Score 61.4;  DB 1;  Length 1511;
  Best Local Similarity   48.0%;  Pred. No. 5e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  ||||  | |||  |   | |  |   |    || |    |
Db    942 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 1001

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db   1002 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 1061

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
             | ||||  |     | ||    | | |||   ||| | ||| | ||||||  |||
Db   1062 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 1121

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db   1122 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 1181

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || |         | ||| |||| |||| ||   |
Db   1182 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 1241

Qy    668 atttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | |||||  ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db   1242 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 1301

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
            |  || ||       |   |  |     |     | | | | |||||| |  ||  |
```

```
Db   1302 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 1361

Qy    788 atataat 794
          | | |||
Db   1362 AAAAAAT 1368


RESULT   8
US-08-544-332-8
; Sequence 8, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
```

```
US-08-544-332-8


  Query Match             5.2%;  Score 61.4;  DB 2;  Length 1511;
  Best Local Similarity   48.0%;  Pred. No. 5e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  ||||  | |||  |   | |  |   |    || |    |
Db    942 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 1001

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db   1002 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 1061

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
           | |||| |     | ||    | | |||   ||| | ||| | ||||||  |||
Db   1062 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 1121

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  ||||
Db   1122 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 1181

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || |          | ||| |||| |||| ||   |
Db   1182 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 1241

Qy    668 atttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db   1242 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 1301

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
           |  || ||       |    |  |      |     | | | | |||||| |  ||  |
Db   1302 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 1361

Qy    788 atataat 794
          | | |||
Db   1362 AAAAAAT 1368


RESULT   9
US-08-852-629-11
; Sequence 11, Application US/08852629
; Patent No. 6106825
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W
;    APPLICANT:  Li, Yi
;    APPLICANT:  Hall, Richard L
;    TITLE OF INVENTION:  ENTOMOPOXVIRUS-VERTEBRATE GENE DELIVERY
;    TITLE OF INVENTION:  VECTOR AND METHOD
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,629
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF-184
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  407-426-7500
```

```
;      TELEFAX:  407-839-8589
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4810 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-852-629-11


  Query Match             5.2%;  Score 61.4;  DB 3;  Length 4810;
  Best Local Similarity   48.0%;  Pred. No. 6.3e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  |||| | ||| |   | |  |   |    || |   |
Db   4235 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 4294

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db   4295 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 4354

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
             | ||||  |     | ||    | | |||   ||| | ||| | ||||||  |||
Db   4355 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 4414

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db   4415 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 4474

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || |          | ||| |||| |||| ||   |
Db   4475 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 4534

Qy    668 atttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db   4535 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 4594

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
            |  || ||      |   |  |       |    | | | | ||||||  |  ||  |
Db   4595 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 4654

Qy    788 atataat 794
          | | |||
Db   4655 AAAAAAT 4661


RESULT  10
US-08-852-629-15
; Sequence 15, Application US/08852629
; Patent No. 6106825
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W
;    APPLICANT:  Li, Yi
;    APPLICANT:  Hall, Richard L
;    TITLE OF INVENTION:  ENTOMOPOXVIRUS-VERTEBRATE GENE DELIVERY
;    TITLE OF INVENTION:  VECTOR AND METHOD
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,629
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF-184
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  407-426-7500
;      TELEFAX:  407-839-8589
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4838 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-852-629-15


  Query Match             5.2%;  Score 61.4;  DB 3;  Length 4838;
  Best Local Similarity   48.0%;  Pred. No. 6.3e-05;
  Matches  205;  Conservative    0;  Mismatches  221;  Indels    1;  Gaps    1;

Qy    369 aattaaaaaagatttagattattaaattattaggttaattcacgggttggctaatcaatt 428
          |||   || | ||| | |  |||| | ||| |   | |  |   |    || |   |
Db   4263 AATATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGAT 4322

Qy    429 attattaattaaaacgatagtatttttgataatttaattaaaatttttattggatttgaat 488
          | || || || |||  ||| |       ||| |  || |  |  | | |   |||| |||
Db   4323 AATAATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAAT 4382

Qy    489 gaactcaattacatcacaaaaaacctaatcaaattaatatcttatgt-gatataatttag 547
             | ||||  |     | ||    | | |||   ||| | ||| | ||||||  |||
Db   4383 ATTATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTAT 4442

Qy    548 aaatataaatgattaacctttaaatctcgagtttctcttataaaaaacacgtataattgg 607
          ||  | ||  |    |    |||   |  | |  | | |   | | |    |  |||||
Db   4443 AACAAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGT 4502

Qy    608 gctagatttaacagctattattcaaactggccaggacaattattaaaattaataattatt 667
          |    || ||| |   || | |  || |          | ||| |||| |||| ||   |
Db   4503 GAATCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTA 4562

Qy    668 atttttttctaataaagcacttcctaattgttaaaatatatgtctaaacactaataataaa 727
          ||  | ||||| ||   | ||  ||| | | | ||| | | ||  |  | ||    | |
Db   4563 ATAATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAG 4622

Qy    728 atttatttgtgtatctttggcagtaggtgagaggtgctgacaaataaattagtgcataaa 787
            |  || ||      |   |  |       |    | | | | ||||||  |  ||  |
Db   4623 TTAAATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTA 4682

Qy    788 atataat 794
          | | |||
Db   4683 AAAAAAT 4689


RESULT  11
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
```

```
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  288:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1241RP
US-08-998-416-288


  Query Match             5.2%;  Score 61.2;  DB 4;  Length 837;
  Best Local Similarity   44.9%;  Pred. No. 4.9e-05;
  Matches  310;  Conservative    0;  Mismatches  378;  Indels    2;  Gaps    2;

Qy      89 tatatatatatatatatgcatgcatgaggaccatggctatgatgaaggttaatagaggta 148
           | || ||||||||  | ||  || | ||  |     |   |     |||  | |  ||
Db       3 TCTAAATATATATAATTTAATTTATAAAGATTAATATAAACTTTTTTATTATAATATTTA 62

Qy     149 gttgtgattgagatatgtccagcactagttttttgttggtgtgatttctcatgatgacgc 208
           | | |      |   | |  | ||  || | |      |    ||||  |
Db      63 AGTATTAAATTATTTAAACTATTATTATCATTATTTAATAAATTAATTATTTGATTATTA 122

Qy     209 gaaaatttttatatatatatataatgaataatatgattgattattctctgtaattttgtga 268
           |  | |||||||  |||   || | | |  | ||| || ||| |   || || | | |
Db     123 ATACTTATTATATAATTATTATATAATTTACTTAATTCATCATTATTAATATTTATAT-A 181

Qy     269 aatagattaaaacagctcaatgtgaggtgaccagttgtcaaatgaccactcgacttgggg 328
           | || |  || |    | ||| |||               |  |    | |  | |  |
Db     182 ATTATAAAAATAATATTTAATATGAATACTATTTAGTCTATGTTCAAATTTTAAATTAGT 241

Qy     329 catggtgatttttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagatt 388
           ||    || ||   |   | | |   ||    ||  ||   | ||| |||||   | |
Db     242 TATTAAAATATTATTAGATATTATTATTTTCTTTAATAAATTATTAAATAGATTATCAAT 301

Qy     389 attaaattattaggttaattcacgggttggctaatcaattattattaattaaaacgatag 448
           | | |||   |  ||| |    |  |     ||| |  |||| |    | | |
Db     302 AATTAATATATTATTTATTAATTGTTTATTAAAATAATATATTTTATTATTATAAAGATT 361

Qy     449 tatttttgataatttaattaaaatttttattggatttgaatgaactcaattacatcacaaa 508
           || |||   ||| |    |||| | |||||    ||  |  | ||   ||     | |
Db     362 TAATTTATTTAAATATTGTAAATTATTATTTTTTATTATAATATCTATTTTTATAAATATT 421

Qy     509 aaacctaatcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttt 568
           |     | | | |||| |   |  | | |||  | |||  | || || | |||     ||
Db     422 ATGTTGATTTATATTATTTAATCTTTTTATAAGAATTATTATTAAAATTAATTTTAACTT 481

Qy     569 aaatctcgagtttctcttataaaaaacacgtataattgggctagatttaacagctattat 628
           |||  ||   ||  |  | |  | |  |  ||  |             |   |  ||||
Db     482 TAATTTCTTATTATTAATTTTATATTATTTAATAAATTATATTTCATTTTATTTATTTA 541

Qy     629 tcaaactggccaggacaattattaaaattaataattattatttttttctaataaagcactt 688
           |  |  |    |    |||||||  |||||||| ||  |  || ||  | |||   |
Db     542 TTTATTTAATTAAATTAATTATT-TAATTAATATTTTATCATTATTTAATTAATTAATAA 600

Qy     689 cctaattgttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggc 748
             || |    | ||| ||  |  ||| ||| |||| |  ||       | |||  |||
Db     601 AATATTATAAAGAATGTAGTTAAAAATACTTATAAAAGGATCCGAACCTATATTATTGTT 660

Qy     749 agtaggtgagaggtgctgacaaataaatta 778
             |  |  | | |   |  |  ||||  ||
Db     661 TATGAGACAAATGCTTTAGCCCATAAGCTA 690


RESULT  12
US-09-316-083-2/c
; Sequence 2, Application US/09316083A
; Patent No. 6280942
; GENERAL INFORMATION:
;  APPLICANT: The Institute of Physical and Chemical Research
;  TITLE OF INVENTION: Endonuclease
;  FILE REFERENCE: PH-651
;  CURRENT APPLICATION NUMBER: US/09/316,083A
;  CURRENT FILING DATE: 1999-05-20
;  EARLIER APPLICATION NUMBER: JP98/141861
;  EARLIER FILING DATE: 1998-05-22
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1431
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1428)
US-09-316-083-2


  Query Match             5.1%;  Score 60;  DB 4;  Length 1431;
  Best Local Similarity   51.6%;  Pred. No. 0.0001;
  Matches  161;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     418 gctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatttttat 477
           |||||  ||||  ||||||  | | | | ||| |||||||  |||  | | |||  |||   |
Db     965 GCTAACCAATAATTATTTTTAATATCCATAATATTTTTAATAGGTAATTTATTATTAATT 906

Qy     478 tggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtga 537
           |  |  | | |  | | |   | || |  |  ||  | || || || |||     | |
Db     905 TTTAAATAATTATAATAAGGGATATGATTATTAATTTGATTAATTTTATAATCACTATAT 846

Qy     538 tataatttagaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacac 597
           ||   ||| ||||||  | | |   | | |||   |  | |   ||   ||  | | |
Db     845 AATTTATTATAAATATTAGTAAAAATCTTTTCTTTAGCTAAATTAGTTTTAATATTAATA 786

Qy     598 gtataattgggctagatttaacagctattattcaaactggcca-ggacaattattaaaat 656
           |   ||        |  |||    |||  || ||| |  |||   |   || |   ||
Db     785 GAGAAAGCATTAACAAAATAAGGATTATCTTTTAAAGTATCCATCCATTTTTCTACTAAA 726

Qy     657 taataattattatttttttctaataaagcacttcctaattgttaaaatatatgtctaaaca 716
             |  ||| ||| || || || |||   |  |  || ||  || |  ||  |  | |  |
Db     725 ATACTATTTTTAATTCTTTTATTAATAAAGATAGTATTTAATACATCATTAGCATGATTA 666

Qy     717 ctaataataaaa 728
           ||||  ||| |
Db     665 GTAATTGTAATA 654
```

```
RESULT  13
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.95238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3770..4013
;      OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;      OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1


  Query Match             5.1%;  Score 60;  DB 2;  Length 6243;
  Best Local Similarity   45.8%;  Pred. No. 0.00013;
  Matches  207;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy    178 ttttgttggtgtgatttctcatgatgacgcgaaaatttttatatatatatataatgaata 237
          |||||  || | | |||| | ||  | |    || ||||||| |   |  || |
Db   2819 TTTTTTATGTTTTAATTTATTATTTTTATTAAAATATTTTAAACTAAATTAAATGATTCT 2878

Qy    238 atatgattgattattctctgtaattttgtgaaatagattaaaacagctcaatgtgaggtg 297
          | | |||  ||| | | |       |   | ||   |  ||  |  |  |  | | |
Db   2879 TTTTAATTTTTTACTATTTCATTCCATAATATATTACTATAATTATTTACAAATAATATT 2938

Qy    298 accagttgtcaaatgaccactcgacttggggcatggtgatttttcaaatcacaactcaat 357
          |   | |  |||    |   || ||      |  | ||||  | || | |    |
Db   2939 TCTTCATTTGTAATATTTAGATGATTTACTAATTTTAGTTTTTATATATTAAATAATTAA 2998

Qy    358 ttgaaaactaaaattaaaaaagatttagattattaaattattaggttaattcacgggttg 417
          |  | || | | || |||||  |       |  |||||||| |   |     | |
Db   2999 TGTATAATTTATATAAAAAATCAAAGGAGCTTATAAATTATGATTATTTCCAAAGATACT 3058

Qy    418 gctaatcaattattattaattaaaacgatagtatttttgataatttaattaaaatttttat 477
            ||  | | || | ||||  ||| ||| | ||| | ||| | |  ||||| ||   |
Db   3059 AAAGATTTAATTTTTTCAATTTTAACAATACTTTTTGTAATATTATGTTTAAATTTAATT 3118

Qy    478 tggatttgaatgaactcaattacatcacaaaaaacctaatcaaattaatatcttatgtga 537
            |||      |    |   |     ||  || | |||| | ||    | | |||| |
Db   3119 GTATTTTTTCATATAATAAAGCCGTTGAAGTAAACCAATCCATTTTCCTTATGATGTTA 3178

Qy    538 tataatttagaaaatataaatgattaacctttaaatctcgagtttctcttataaaaaacac 597
          |          ||  |  || || |    || | | |     || | ||  |||||||
Db   3179 TTATTAAATTTAAGTTTTATAATAATATCTTTATTATATTTATTGTTTTTAAAAAAACTA 3238

Qy    598 gtataattgggctagatttaacagctattatt 629
          ||  ||||       | | | |  |||| ||
Db   3239 GTGAAATTTCCGGCTTTATTAAACTTATTTTT 3270


RESULT  14
US-08-998-416-1137/c
; Sequence 1137, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1137:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1692RP
US-08-998-416-1137


  Query Match             5.0%;  Score 58.8;  DB 4;  Length 636;
  Best Local Similarity   48.3%;  Pred. No. 0.00016;
  Matches  224;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    2;

Qy    339 tttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaatt-- 396
          ||   ||| | || | ||  | |||| |||  |||| ||  |||   || |  | |||
Db    498 TTAATAATAAGAAATTAAAGTTAAAATTAATTTTAATAATAATTCTTATAAAAAGATTAA 439
```

```
Qy    397 attaggttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttg 456
          || |  | |||  ||    |   ||   || ||    | || | ||  ||| || |||
Db    438 ATAATATAAATCAACATAATATTTATAAAAATAGATATTATAATAAAAATAATAATTTAC 379

Qy    457 ataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaaaacctaa 516
          |  |||||| |||| |     |  ||   ||| || |  ||||  | |  ||| |  |||
Db    378 AATATTTAAATAAATTAAATCTTTATAATAATAAAATATATTATTTTAATAAACAATTAA 319

Qy    517 tcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcg 576
          | ||    ||||   | |    |  |   ||||        |     |||  |    -
Db    318 TAAATAATATATTAATTATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATAT 259

Qy    577 agtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaactg 636
              |    | | || |||   | |||       | ||  |  |  || ||||||  | |
Db    258 CTAATAATATTTTAATAACTAATTTAAAATTTGAACATAGACTAAATAGTATTCATATTA 199

Qy    637 gccaggacaattattaaaat-taataattattatttttttctaataaagcacttcctaatt 695
            |  |   |||| |  || |||  |||| || |  |      |   | |  ||||
Db    198 AATATTATTTTTATAATTATATAAATATTAATAATGATGAATTAAGTAAATTATATAATA 139

Qy    696 gttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggt 755
           ||| |  ||| || | || | | | |  |  | ||  ||   ||    |  | ||| |
Db    138 ATTATATAATAAGTATTAATAATCAAATAATTAATTTATTAAATAATGATAATAATAGTT 79

Qy    756 gagaggtgctgacaaataaattagtgcataaaatataatggatt 799
            | |     | | |  |||||      |||||| |   |  |||
Db     78 TAAATAATTTAATACTTAAATATTATAATAAAAAAGTTTATATT 35


RESULT 15
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  288:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1241RP
US-08-998-416-288


  Query Match             5.0%;  Score 58.8;  DB 4;  Length 837;
  Best Local Similarity   48.3%;  Pred. No. 0.00016;
  Matches  224;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    2;

Qy    339 tttcaaatcacaactcaatttgaaaactaaaattaaaaaagatttagattattaaatt-- 396
          ||   ||| | || | ||  | |||| |||  |||| ||  |||   || |  | |||
Db    498 TTAATAATAAGAAATTAAAGTTAAAATTAATTTTAATAATAATTCTTATAAAAAGATTAA 439

Qy    397 attaggttaattcacgggttggctaatcaattattattaattaaaacgatagtatttttg 456
          || |  | |||  ||    |   ||   || ||    | || | ||  ||| || |||
Db    438 ATAATATAAATCAACATAATATTTATAAAAATAGATATTATAATAAAAATAATAATTTAC 379

Qy    457 ataatttaattaaaattttattggatttgaatgaactcaattacatcacaaaaaacctaa 516
          |  |||||| |||| |     |  ||   ||| || |  ||||  | |  ||| |  |||
Db    378 AATATTTAAATAAATTAAATCTTTATAATAATAAAATATATTATTTTAATAAACAATTAA 319

Qy    517 tcaaattaatatcttatgtgatataatttagaaatataaatgattaacctttaaatctcg 576
          | ||    ||||   | |    |  |   ||||        |     |||  |
Db    318 TAAATAATATATTAATTATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATAT 259

Qy    577 agtttctcttataaaaaacacgtataattgggctagatttaacagctattattcaaactg 636
              |    | | || |||   | |||       | ||  |  |  || ||||||  | |
Db    258 CTAATAATATTTTAATAACTAATTTAAAATTTGAACATAGACTAAATAGTATTCATATTA 199

Qy    637 gccaggacaattattaaaat-taataattattatttttttctaataaagcacttcctaatt 695
            |  |   |||| |  || |||  |||| || |  |      |   | |  ||||
Db    198 AATATTATTTTTATAATTATATAAATATTAATAATGATGAATTAAGTAAATTATATAATA 139

Qy    696 gttaaaatatatgtctaaacactaataataaaatttatttgtgtatctttggcagtaggt 755
           ||| |  ||| || | || | | | |  |  | ||  ||   ||    |  | ||| |
Db    138 ATTATATAATAAGTATTAATAATCAAATAATTAATTTATTAAATAATGATAATAATAGTT 79

Qy    756 gagaggtgctgacaaataaattagtgcataaaatataatggatt 799
            | |     | | |  |||||      |||||| |   |  |||
Db     78 TAAATAATTTAATACTTAAATATTATAATAAAAAAGTTTATATT 35


Search completed: July 30, 2002, 11:11:03
Job time: 3987 sec
```

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 21:58:04 ; Search time 1921.77 Seconds
(without alignments)
174.227 Million cell updates/sec

Title: US-09-530-663B-15
Perfect score: 16
Sequence: 1 cctttcaccaaccccc 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16	100.0	1172	8	AF041051	AF041051 Populus t
	2	16	100.0	4397	8	AY043494	AY043494 Populus t
	3	16	100.0	40114	3	AF067942	AF067942 Caenorhab
c	4	16	100.0	173705	2	AC012256	AC012256 Homo sapi
	5	16	100.0	177050	10	AC084390	AC084390 Mus muscu
c	6	16	100.0	221647	10	AL591003	AL591003 Mouse DNA
c	7	15	93.8	3545	8	SCYOL075C	Z74817 S.cerevisia
c	8	15	93.8	4423	1	HPU86609	U86609 Helicobacte
	9	15	93.8	17880	1	AE000625	AE000625 Helicobac
	10	15	93.8	26236	2	AC107593	AC107593 Rattus no
	11	15	93.8	26458	9	AC005942	AC005942 Homo sapi
	12	15	93.8	30208	9	HS212J10	AL034401 Human DNA
	13	15	93.8	46363	2	AC099662	AC099662 Rattus no
	14	15	93.8	62490	2	AC110265	AC110265 Mus muscu
	15	15	93.8	76122	2	AC098003	AC098003 Rattus no
c	16	15	93.8	78379	8	AB006696	AB006696 Arabidops
	17	15	93.8	84592	2	AC016403	AC016403 Homo sapi
c	18	15	93.8	91894	9	AC005739	AC005739 Homo sapi
c	19	15	93.8	92134	2	AC018406	AC018406 Homo sapi
	20	15	93.8	94908	2	AC099247	AC099247 Rattus no
c	21	15	93.8	126062	2	AC011363	AC011363 Homo sapi
c	22	15	93.8	131312	9	AC008456	AC008456 Homo sapi
	23	15	93.8	133563	2	AC099481	AC099481 Homo sapi
	24	15	93.8	133589	2	AC103534	AC103534 Rattus no
	25	15	93.8	134078	9	AC068515	AC068515 Homo sapi
	26	15	93.8	137330	2	AC092346	AC092346 Homo sapi
	27	15	93.8	150157	2	AC103150	AC103150 Rattus no
	28	15	93.8	150288	2	AC015572	AC015572 Homo sapi
	29	15	93.8	150288	9	AC015574	AC015574 Homo sapi
	30	15	93.8	151872	2	AC084250	AC084250 Homo sapi
c	31	15	93.8	152743	2	AC103511	AC103511 Rattus no
	32	15	93.8	153647	2	AC079992	AC079992 Homo sapi
c	33	15	93.8	157487	2	AC095944	AC095944 Rattus no
	34	15	93.8	157656	2	AC104124	AC104124 Homo sapi
	35	15	93.8	158001	2	AC104365	AC104365 Homo sapi
	36	15	93.8	159150	2	AC022863	AC022863 Homo sapi
c	37	15	93.8	161775	9	AC099512	AC099512 Homo sapi
	38	15	93.8	164958	2	AC094338	AC094338 Rattus no
	39	15	93.8	167438	2	AC095832	AC095832 Rattus no
c	40	15	93.8	171949	9	AC012531	AC012531 Homo sapi
c	41	15	93.8	172403	2	AC017061	AC017061 Homo sapi
c	42	15	93.8	177512	2	AC099364	AC099364 Rattus no
c	43	15	93.8	177968	30	AC022736	Ac022736 Homo sapi
c	44	15	93.8	180638	9	AC068763	AC068763 Homo sapi
c	45	15	93.8	181127	2	AC068203	AC068203 Homo sapi

ALIGNMENTS

RESULT 1
AF041051
LOCUS AF041051 1172 bp DNA linear PLN 26-JUN-1998
DEFINITION Populus tremuloides clone Pt4CL1p 4-coumarate:CoA ligase gene, promoter region.
ACCESSION AF041051
VERSION AF041051.1 GI:3258638
KEYWORDS .
SOURCE quaking aspen.
ORGANISM Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 1172)
AUTHORS Hu,W.J., Kawaoka,A., Tsai,C.J., Lung,J., Osakabe,K., Ebinuma,H. and Chiang,V.L.
TITLE Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (Populus tremuloides)
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (9), 5407-5412 (1998)
MEDLINE 98226828
REFERENCE 2 (bases 1 to 1172)
AUTHORS Hu,W.-J. and Chiang,V.L.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1998) Plant Biotechnology Research Center, School
            of Forestry and Wood Products, Michigan Technological University,
            1400 Townsend Drive, Houghton, MI 49931, USA
FEATURES             Location/Qualifiers
     source          1. .1172
                     /organism="Populus tremuloides"
                     /db_xref="taxon:3693"
                     /clone="Pt4CL1p"
                     /tissue_type="leaf"
     gene            1. .>1172
                     /gene="4-coumarate:CoA ligase"
     promoter        1. .1172
                     /gene="4-coumarate:CoA ligase"
BASE COUNT      399 a    224 c    181 g    368 t
ORIGIN


  Query Match             100.0%;  Score 16;  DB 8;  Length 1172;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db   1018 CCTTTCACCAACCCCC 1033


RESULT   2
AY043494
LOCUS       AY043494                4397 bp    DNA     linear   PLN 17-SEP-2001
DEFINITION  Populus tomentosa 4-coumarate:CoA ligase gene, complete cds.
ACCESSION   AY043494
VERSION     AY043494.1  GI:15636676
KEYWORDS    .
SOURCE      Populus tomentosa.
  ORGANISM  Populus tomentosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE   1  (bases 1 to 4397)
  AUTHORS   Lu,H., Zeng,Q. and Jiang,X.
  TITLE     Genomic DNA sequence of 4-coumarate:CoA ligase gene from Populus
            tomentosa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4397)
  AUTHORS   Lu,H., Zeng,Q. and Jiang,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2001) Biology, Beijing Forestry University,
            Beijing Qinghua East Road No. 35, Beijing 100081, China
FEATURES             Location/Qualifiers
     source          1. .4397
                     /organism="Populus tomentosa"
                     /db_xref="taxon:118781"
     promoter        1. .1136
     mRNA            join(1137. .2136,3321. .3519,3644. .3857,3957. .4059,
                     4174. .4397)
                     /product="4-coumarate:CoA ligase"
     CDS             join(1153. .2136,3321. .3519,3644. .3857,3957. .4059,
                     4174. .4284)
                     /codon_start=1
                     /product="4-coumarate:CoA ligase"
                     /protein_id="AAL02144.1"
                     /db_xref="GI:15636677"
                     /translation="MNPQEEFIFRSKLPDIYIPKNLPLHSYVLENLSNHSSKPCLING
                     ANGDVYTYADVELTARRVASGLNKIGIQQGDVIMLFLPSSPEFVLAFLGASHRGAIIT
                     AANPFSTPAELAKHAKASRAKLLITQACYYEKVKDFARESDVKVMCVDSAPDGCLHFS
                     ELTQADENEAPQVDISPDDVVALPYSSGTTGLPKGVMLTHKGLITSVAQQVDGDNPNL
                     YFHSEDVILCVLPMFHIYALNSIMLCGLRVGASILIMPKFEIGSLLGLIEKYKVSIAP
                     VVPPVMMSIAKSPDLDKHDLSSLRMIKSGGAPLGKELEDTVRAKFPQARLGQGYGMTE
                     AGPVLAMCLAFAKEPFDIKPGACGTVVRNAEMKIVDPETGASLPRNQPGEICIRGDQI
                     MKGYLNDPEATSRTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPAELE
                     ALLIAHPEISDAAVVGLKDEDAGEVPVAFVVKSEKSQATEDEIKQYISKQVIFYKRIK
                     RVFFIEAIPKAPSGKILRKNLKEKLAGI"

BASE COUNT     1394 a    823 c    819 g    1361 t
ORIGIN


  Query Match             100.0%;  Score 16;  DB 8;  Length 4397;
  Best Local Similarity   100.0%;  Pred. No. 95;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db   1021 CCTTTCACCAACCCCC 1036


RESULT   3
AF067942
LOCUS       AF067942                40114 bp    DNA     linear   INV 05-OCT-2001
DEFINITION  Caenorhabditis elegans cosmid ZK6, complete sequence.
ACCESSION   AF067942
VERSION     AF067942.1  GI:3165573
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40114)
  AUTHORS   The C. elegans Sequencing Consortium.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
REFERENCE   2  (bases 1 to 40114)
  AUTHORS   Wu,X.
  TITLE     The sequence of C. elegans cosmid ZK6
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 40114)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: rw@nematode.wustl.edu and jes@sanger.ac.uk
```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y39D8B;3' cosmid is Y39D8A. Actual start of this cosmid is at base position 1 of CELZK6; actual end is at 40114 of CELZK6. The orientation of this cosmid is unknown. .

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers

source 1. .40114
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"
/clone="ZK6"

gene 2105. .4708
/gene="ZK6.8"

CDS join(2105. .2307,3036. .3223,3287. .3515,3586. .3905, 3956. .4127,4306. .4441,4490. .4708)
/gene="ZK6.8"
/codon_start=1
/product="Hypothetical protein ZK6.8"
/protein_id="AAG45570.1"
/db_xref="GI:12019644"
/translation="MSEESLPPSSGSKMSPRAFELLCVVMLGFGHLCIMTGCDSQAFI
LESVINSIHERDPARINSHAGYYGQATCYLAFVFTCLVSPTFLYATSAKTTLIIAAAC
FTSFPLGFLYTNQYYYFFSAALNGVGFALYYTGNGGYLTSHSTRKTIESNVSISWAIG
SSCMIVGAGIIALITFLTAGQGSEVAMDLANATVTQHFERRFSDTEIYLLFSVFAAIS
FVGCITFALLPSNDIGNCIESSKKIVAFRDGIALMYRAFRSPKMIVLIPTFVLTGVHT
SFWVSIYPTTLTFNSHLSAMIYLPAIYSFGVGLGETTMGLLISFCSKRIKNFGMRPTM
FIGCFLTCLYCALVVISTPPTAPMAPTSEKPLLFQPTRYLVFIIALIGGMSDCCLCSV
RSVVCAIAMPTRRNQAFSVSKFYQAIGCCVIFFISPLLNIYYYVIGIPILCIIASVCF
FFETRRIKQMEKSLTNMELDQAQQRRRSSKYDTLDEEF"

gene 5937. .7881
/gene="ZK6.7"

CDS join(5937. .6023,6070. .6169,6669. .7138,7228. .7527, 7589. .7687)
/gene="ZK6.7"
/note="coded for by the following C. elegans cDNAs: yk788c11.3, yk788c11.5"
/codon_start=1
/product="Hypothetical protein ZK6.7b"
/protein_id="AAL11486.1"
/db_xref="GI:15967144"
/translation="MWRFAVFLAAFFVQDVVGSHGDPELHMTTPQIIERWGYPAMIYT
VATDDGYILEMHRIPFGKTNVTWPNGKRPVVFMQHGLLCASSDWVVNLPDQSAGFLFA
DAGFDVWLGNMRGNTYSMKHKDLKPSHSAFWDWSWDEMATYDLNAMINHVLEVTGQDS
VYYMGHSQGTLTMFSHLSKDDGSFAKKIKKFFALAPIGSVKHIKGFLSFFANYFSLEF
DGWFDIFGAGEFLPNNWAMKLAAKDICGGLKVEADLCDNVLFLIAGPESDQWNQTRVP
VYATHDPAGTSTQNIVHWMQMVHHGGVPAYDWGTKTNKKKYGQSTTLPLSREPRFTST
GVMPIGWLIRLTCLTTY"

CDS join(5937. .6023,6070. .6169,6669. .7138,7228. .7527, 7576. .7713,7765. .7881)
/gene="ZK6.7"
/note="similar to lysosomal acid lipases (SW:P38571); coded for by the following C. elegans cDNAs: yk263b10.3, yk263b10.5, yk308a8.5, yk552a3.3, yk552a3.5, yk722d9.5, yk722d9.3, yk733a11.5, yk733g10.5, yk736f6.5, yk736f6.3, yk733g10.3, yk851d01.3, yk851d01.5"
/codon_start=1
/product="Hypothetical protein ZK6.7a"
/protein_id="AAG45574.1"
/db_xref="GI:12019648"
/translation="MWRFAVFLAAFFVQDVVGSHGDPELHMTTPQIIERWGYPAMIYT
VATDDGYILEMHRIPFGKTNVTWPNGKRPVVFMQHGLLCASSDWVVNLPDQSAGFLFA
DAGFDVWLGNMRGNTYSMKHKDLKPSHSAFWDWSWDEMATYDLNAMINHVLEVTGQDS
VYYMGHSQGTLTMFSHLSKDDGSFAKKIKKFFALAPIGSVKHIKGFLSFFANYFSLEF
DGWFDIFGAGEFLPNNWAMKLAAKDICGGLKVEADLCDNVLFLIAGPESDQWNQTRVP
VYATHDPAGTSTQNIVHWMQMVHHGGVPAYDWGTKTNKKKYGQANPPEYDFTAIKGTK
IYLYWSDADWLADTPDVPDYLLTRLNPAIVAQNNHLPDYNHLDFTWGLRAPDDIYRPA
IKLCTDDYLGK"

gene 9204. .11716
/gene="ZK6.6"

CDS join(9204. .9343,9554. .9741,9787. .9883,9943. .10074, 10127. .10257,10841. .11029,11081. .11256,11314. .11352, 11501. .11716)
/gene="ZK6.6"
/codon_start=1
/product="Hypothetical protein ZK6.6"
/protein_id="AAG45573.1"
/db_xref="GI:12019647"
/translation="MDKFCDETFYKLKKSYGSYKKILFLEKNFAKLQRNFNVVFYTYC
DLYQAVCYLAYVTTCLFSPSFLYATSAKTTLLISSICFTSFPLGFLFTNSYYYYFSSA
LNGVGFALYYTGNGGYITSHSTRQTIESNVSLSWSFGCCCMIVGSVIMAVITTVTQDQ
IPQLIDAVNVTIGGHGVERRFGDGEIYLLFSVFAAISSMAIFTFMLMPSQDVSNCIEP
SDKVVSFKDGMKLMAETLKSSKMFKLAPTFLLTGMYTAFWVSIYPTSLTFNMHNSKMI
YLPAIYGFGVGVGETIMGIIISTLSKRIKDFGLKPTMLIGSVLTTVFCFVALLSTPFD
ATVTPSHEQPLLFQPRRPSYSISDCNNRRNSLGSCFIFFLSPFLNLYHYVIGIPLLCV
LSCILFFRQARQTQVMERKLTQELEESEKRRMAKELEAQMQKI"

gene 12832. .14865
/gene="ZK6.5"

CDS join(12832. .12894,13721. .13817,13864. .14452,14605. .14686, 14731. .14865)
/gene="ZK6.5"
/note="contains similarity to C4-type zinc finger (Pfam; zf-C4.hmm, score: 37.95)"
/codon_start=1
/product="Hypothetical protein ZK6.5"
/protein_id="AAG45580.1"
/db_xref="GI:12019654"
/translation="MLASNQTCAVCERFTTEFNYGVPSCNACKIFFRRLITRTAPVKQ
CYIGEHCFTKSPITKKCTFCRFQKCIQVGMTLPSYLHFGELTKEKCIDSTIQKLILME
AHRKDLMANYYTSYLNPTIDEVIRLNKVEYTRKSQNHQMSFYNWAFHCCLVTVDFMKK
FSFVNLLRFEDQKNLMKEFYIKLTVLINSKQSMSCGKEGMTFPDGSDVLPPTSSEWGI
SKISQNLENKVRCRLIGRLSELRITDEEYLLINFKNFQKFSKILITFSSVFKAFYNFP
RTELSKIVSRKIEKHFQSTKILQIFPEVSSIFQKNLKFSTTFQNS"

gene 17012. .18941
/gene="ZK6.4"

CDS join(17012. .17074,17511. .17607,17658. .18272,18679. .18941)
/gene="ZK6.4"
/note="contains similarity to C4-type zinc finger (Pfam; zf-C4.hmm, score: 34.58) and to the ligand-binding domain of nuclear hormone receptors (Pfam: hormone_rec.hmm, score: 38.56)"
/codon_start=1
/product="Hypothetical protein ZK6.4"
/protein_id="AAG45579.1"
/db_xref="GI:12019653"
/translation="MVATNQKCAVCERFTTEFNYGVPSCNACKIFFRRLITRTTPVKK
CYLGENCFTKTPITQKCVFCRFQKCIQVGMTLPSYLHFGEQTKEKCLDSTIQNLMLLD
AHREDLMANYHTSYLDPTIDEVIRINKVDYIRKSQSHQLNFYNWAFHCSLVTVDFMKK
FPFVNLLRFEDQKYLLKEFYIKLIVFISSKRSMSRGKECMTFPDGSDVLPPASSEWHI
SKISPNLENKIRCRLIGRLSELRITDEEYLLMCVLIFCNPSLSQLSENGRRLLTSYQH
LYSSALLQYCLLTYEKAGPARFAELLGICHVIGVQYDDAIHYNVLVQLTKARLDVKQL
IKDGMDAAFCN"

gene 20442. .21787
/gene="ZK6.3"

CDS join(20442. .20615,20667. .20801,21151. .21222,21444. .21631,

Query Match 100.0%; Score 16; DB 3; Length 40114;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db  15000 CCTTTCACCAACCCCC 15015


RESULT  4
AC012256/c
LOCUS       AC012256              173705 bp    DNA     linear   HTG 18-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-135G15 map 8, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC012256
VERSION     AC012256.5  GI:13376921
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173705)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-135G15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173705)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 18, 2001 this sequence version replaced gi:13357415.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3739
                Center clone name: 135_G_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 69% of reads
                Sequencing vector: Plasmid; n/a; 31% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 163867 bases at least Q40
                Consensus quality: 168063 bases at least Q30
                Consensus quality: 169817 bases at least Q20
                Insert size: 172005; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1853: contig of 1853 bp in length
            *     1854 1953: gap of      100 bp
            *     1954     3364: contig of 1411 bp in length
            *     3365 3464: gap of      100 bp
            *     3465     4706: contig of 1242 bp in length
            *     4707 4806: gap of      100 bp
            *     4807    27066: contig of 22260 bp in length
            *    27067 27166: gap of      100 bp
            *    27167    30445: contig of 3279 bp in length
            *    30446 30545: gap of      100 bp
            *    30546    34289: contig of 3744 bp in length
            *    34290 34389: gap of      100 bp
            *    34390    40744: contig of 6355 bp in length
            *    40745 40844: gap of      100 bp
            *    40845    46688: contig of 5844 bp in length
            *    46689 46788: gap of      100 bp
            *    46789    55147: contig of 8359 bp in length
            *    55148 55247: gap of      100 bp
            *    55248    65028: contig of 9781 bp in length
            *    65029 65128: gap of      100 bp
            *    65129    79378: contig of 14250 bp in length
            *    79379 79478: gap of      100 bp
            *    79479    93773: contig of 14295 bp in length
            *    93774 93873: gap of      100 bp
            *    93874   107619: contig of 13746 bp in length
            *   107620 107719: gap of      100 bp
            *   107720   122838: contig of 15119 bp in length
            *   122839 122938: gap of      100 bp
            *   122939   136635: contig of 13697 bp in length
            *   136636 136735: gap of      100 bp
            *   136736   152397: contig of 15662 bp in length
            *   152398 152497: gap of      100 bp
            *   152498   173172: contig of 20675 bp in length
            *   173173 173272: gap of      100 bp
            *   173273   173705: contig of 433 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173705
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-135G15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1853
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1954. .3364
                     /note="assembly_fragment"
     misc_feature    3465. .4706
                     /note="assembly_fragment"
     misc_feature    4807. .27066
                     /note="assembly_fragment"
     misc_feature    27167. .30445
                     /note="assembly_fragment"
     misc_feature    30546. .34289
                     /note="assembly_fragment"
     misc_feature    34390. .40744
                     /note="assembly_fragment"
     misc_feature    40845. .46688
                     /note="assembly_fragment"
     misc_feature    46789. .55147
                     /note="assembly_fragment"
     misc_feature    55248. .65028
                     /note="assembly_fragment"
     misc_feature    65129. .79378
                     /note="assembly_fragment"
     misc_feature    79479. .93773
                     /note="assembly_fragment"
     misc_feature    93874. .107619
                     /note="assembly_fragment"
     misc_feature    107720. .122838
                     /note="assembly_fragment"
     misc_feature    122939. .136635
                     /note="assembly_fragment"
```

```
     misc_feature    136736. .152397
                     /note="assembly_fragment"
     misc_feature    152498. .173172
                     /note="assembly_fragment"
     misc_feature    173273. .173705
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54075 a  32056 c  33141 g  52665 t   1768 others
ORIGIN


  Query Match             100.0%;  Score 16;  DB 2;  Length 173705;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db  18624 CCTTTCACCAACCCCC 18609


RESULT   5
AC084390
LOCUS       AC084390              177050 bp    DNA     linear   ROD 26-SEP-2001
DEFINITION  Mus musculus clone RP23-285H9, complete sequence.
ACCESSION   AC084390
VERSION     AC084390.1  GI:15778823
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 177050)
  AUTHORS   McPherson,J.D.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177050)
  AUTHORS   McPherson,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2001) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, Missouri 63108, USA
COMMENT     Center project name: M_BA0285H09.
FEATURES             Location/Qualifiers
     source          1. .177050
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="m"
                     /clone="RP23-285H9"
BASE COUNT    47756 a  40786 c  41518 g  46990 t
ORIGIN


  Query Match             100.0%;  Score 16;  DB 10;  Length 177050;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db  17118 CCTTTCACCAACCCCC 17133


RESULT   6
AL591003/c
LOCUS       AL591003              221647 bp    DNA     linear   ROD 29-NOV-2001
DEFINITION  Mouse DNA sequence from clone RP23-354I24 on chromosome 13,
            complete sequence.
ACCESSION   AL591003
VERSION     AL591003.16  GI:16605732
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 221647)
  AUTHORS   Kay,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 3, 2001 this sequence version replaced gi:16030192.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-354I24 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-354I24.
FEATURES             Location/Qualifiers
     source          1. .221647
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-354I24"
                     /clone_lib="RPCI-23"
     misc_feature    50173. .50333
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
BASE COUNT    52058 a  59303 c  59031 g  51255 t
ORIGIN


  Query Match             100.0%;  Score 16;  DB 10;  Length 221647;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db 152713 CCTTTCACCAACCCCC 152698


RESULT   7
SCYOL075C/c
LOCUS       SCYOL075C              3545 bp    DNA     linear   PLN 05-AUG-1997
DEFINITION  S.cerevisiae chromosome XV reading frame ORF YOL075c.
ACCESSION   Z74817 Y13140
VERSION     Z74817.1  GI:1419904
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3545)
  AUTHORS   Alexandraki,D., Katsoulou,C. and Tzermia,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3545)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XV sequencing project. MIPS at the
```

```
          Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
          Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .3545
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XV"
     CDS             complement(244. .>3545)
                     /note="ORF YOL075c"
                     /codon_start=3
                     /protein_id="CAA99085.1"
                     /db_xref="GI:2315260"
                     /translation="IGTQMISNPSIMFLDEPTTGLDAYSAFLVIKTLKKLAKEDGRTF
                     IMSIHQPRSDILFLLDQVCILSKGNVVYCDKMDNTIPYFESIGYHVPQLVNPADYFID
                     LSSVDSRSDKEEAATQSRLNSLIDHWHDYERTHLQLQAESYISNATEIQIQNMTTRLP
                     FWKQVTVLTRRNFKLNFSDYVTLISTFAEPLIIGTVCGWIYYKPDKSSIGGLRTTTAC
                     LYASTILQCYLYLLFDTYRLCEQDIALYDRERAEGSVTPLAFIVARKISLFLSDDFAM
                     TMIFVSITYFMFGLEADARKFFYQFAVVFLCQLSCSGLSMLSVAVSRDFSKASLVGNM
                     TFTVLSMGCGFFVNAKVMPVYVRWIKYIAFTWYSFGTLMSSTFTNSYCTTDNLDECLG
                     NQILEVYGFPRNWITVPAVVLLCWSVGYFVVGAIILYLHKIDITLQNEVKSKQKKIKK
                     KSPTGMKPEIQLLDDVYHQKDLEAEKGKNIHITIKLEDIDLRVIFSAPFSNWKEGNFH
                     HETKEILQSVNAIFKPGMINAIMGPSGSGKSSLLNLISGRLKSSVFAKFDTSGSIMFN
                     DIQVSELMFKNVCSYVSQDDDHLLAALTVKETLKYAAALRLHHLTEAERMERTDNLIR
                     SLGLKHCENNIIGNEFVKGISGGEKRRVTMGVQLLNDPPILLLDEPTSGLDSFTSATI
                     LEILEKLCREQGKTIIITIHQPRSELFKRFGNVLLLAKSGRTAFNGSPDEMIAYFTEL
                     GYNCPSFTNVADFFLDLISVNTQNEQNEISSRARVEKILSAWKANMDNESLSPTPISE
                     KQQYSQESFFTEYSEFVRKPANLVLAYIVNVKRQFTTTRRSFDSLMARIAQIPGLGVI
                     FALFFAPVKHNYTSISNRLGLAQESTALYFVGMLGNLACYPTERDYFYEEYNDNVYGI
                     APFFLAYMTLELPLSALASVLYAVFTVLACGLPRTAGNFFATVYCSFIVTCCGERLGI
                     MTNTFFERPGFVVNCISIILSIGTQMSGLMSLGMSRVLKGFNYLNPVGYTSMIIINFA
                     FPGNLKLTCEDGGKNSDGTCEFANGHDVLVSYGLVRNTQKYLGIIVCVAIIYRLIAFF
                     ILKAKLEWIKW"
BASE COUNT     1156 a    707 c    585 g   1097 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 8;  Length 3545;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          ||||||||||||||||
Db   1174 CCTTTCACCAACCCC 1160


RESULT   8
HPU86609/c
LOCUS       HPU86609                4423 bp    DNA     linear   BCT 31-AUG-1999
DEFINITION  Helicobacter pylori ribosomal protein L1 homolog gene, partial cds,
            50S ribosomal protein L7/L12 gene, complete cds, and DNA-directed
            RNA polymerase homolog gene, partial cds.
ACCESSION   U86609
VERSION     U86609.1  GI:1840149
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 4423)
  AUTHORS   Hocking,D., Webb,E., Radcliff,F., Rothel,L., Taylor,S.,
            Pinczower,G., Kapouleas,C., Braley,H., Lee,A. and Doidge,C.
  TITLE     Isolation of recombinant protective Helicobacter pylori antigens
  JOURNAL   Infect. Immun. 67 (9), 4713-4719 (1999)
  MEDLINE   99386867
   PUBMED   10456921
REFERENCE   2  (bases 1 to 4423)
  AUTHORS   Hocking,D., Rothel,L., Doidge,C., Radcliff,F., Lee,A. and Webb,E.
  TITLE     Recombinant Helicobacter pylori ribosomal protein of 13kDa mass,
            predicted to be a 50S ribosomal protein L7/L12
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 4423)
  AUTHORS   Hocking,D., Rothel,L., Doidge,C., Radcliff,F., Lee,A. and Webb,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1997) Microbiology Research, CSL Ltd., 45 Poplar
            Rd, Parkville, Victoria 3052, Australia
FEATURES             Location/Qualifiers
     source          1. .4423
                     /organism="Helicobacter pylori"
                     /strain="Hp921023"
                     /db_xref="taxon:210"
     CDS             <1. .538
                     /note="similar to bacterial 50S ribosomal protein L1"
                     /codon_start=2
                     /transl_table=11
                     /product="ribosomal protein L1 homolog"
                     /protein_id="AAB47277.1"
                     /db_xref="GI:1840150"
                     /translation="DQMVRGAVVLPHGTGKKVRVAVFAKDIKQDEAKNAGADVVGGDD
                     LAEEIKNGRIDFDMVIATPDMMAVVGKVGRILGPKGLMPNPKTGTVTMDIAKAVTNAK
                     SGQVNFRVDKKGNVHAPIGKASFPEEKIKENMLELVKTINRLKPSSAKGKYIRNAALS
                     LTMSPSVSLDAQELMDIK"
     CDS             1186. .1563
                     /note="13 kDa protein"
                     /codon_start=1
                     /transl_table=11
                     /product="50S ribosomal protein L7/L12"
                     /protein_id="AAB47278.1"
                     /db_xref="GI:1840151"
                     /translation="MAISKEEVLEYIGSLSVLELSELVKMFEEKFGVSATPTVVAGAA
                     VAGGAAAESEEKTEFNVILADSGAEKIKVIKVVREITGLGLKEAKDATEKTPHVLKEG
                     VNKEEAETIKKKLEEVGAKVEVK"
     CDS             1783. .>4423
                     /note="similar to C. jejuni DNA-directed RNA polymerase
                     and to E. coli DNA-directed RNA polymerase beta chain"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-directed RNA polymerase homolog"
                     /protein_id="AAB47279.1"
                     /db_xref="GI:1840152"
                     /translation="MSKKIPLKNRLRADFTKTPTDLEVPNLLLLQRDSYDSFLYSKDG
                     KESGIEKVFKSIFPIQDEHNRITLEYAGCEFGKSKYTVREAMERGITYSIPLKIKVRL
                     ILWEKDTRSGEKNGIKDIKEQSIFIREIPLMTERTSFIINGVERVVVNQLHRSPGVIF
                     KEEESSTSLNKLIYTGQIIPDRGSWLYFEYDSKDVLYARINKRRKVPVTILFRAMDYQ
                     KQDIIKMFYPLVKVRYENDKYLIPFASLDANQRMEFDLKDPQGKIILLAGKKLTSRKI
                     KELKENHLEWVEYPMDILINRHLAEPVMVGKEVLLDMLTQLDKNKLEKIHDLGVQEFV
                     IINDLALGHDASIIHSFLADYESLKLLKQTEKIDDENALAAIRIHKVMKPGDPVTTEV
                     AKQFVKKLFFDPERYDLTMVGRMKMNHKLGLHVPDYITTLTHEDIITTVKYLMKIKNN
                     QGKIDDRDHLGNRRIRAVGELLANELHSGLVKMQKTIKDKLTTMSGAFDSLMPHDLVN
                     SKMITSTIMEFFMGGQLSQFMDQTNPLSEVTHKRRLSALGEGGLVKDRVGFEARDVHP
                     THYGRICPIETPEGQNIGLINTLSTFTRVNDLGFIEAPYKKVVDGKVVGETIYLTAIQ
                     EDSHIIAPASTPIDEEGNILGDLIETRVEGEIVLNEKSKVTLMDLSSSMLVGVAASLI
                     PFLEHDDANRALMGTNMQRQAVPLLRSDAPIVGTGIEKIIARDSWGAIKANRAGVVEK
                     IDSKNIYILGEGKEEAYIDAYSLQKNLRTNQNTSFNQVPIVKVGDKVEAGQIIADGPS
                     MDRGELALGKNVRVAFMPWNGYNFEDAIVVSERITKDDIFTSTHIYEKEVDARELKHG
                     VEEFTADIPDVKEEALAHLDESGI"
BASE COUNT     1395 a    786 c   1057 g   1185 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 1;  Length 4423;
  Best Local Similarity   100.0%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ctttcaccaacccccc 16
          ||||||||||||||||
Db   3445 CTTTCACCAACCCCC 3431


RESULT   9
AE000625
LOCUS       AE000625                17880 bp    DNA     linear   BCT 06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 103 of 134 of the complete
            genome.
ACCESSION   AE000625 AE000511
VERSION     AE000625.1  GI:2314349
KEYWORDS    .
```

SOURCE Helicobacter pylori 26695.
ORGANISM Helicobacter pylori 26695
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 17880)
AUTHORS Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G., Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R., Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
TITLE The complete genome sequence of the gastric pathogen Helicobacter pylori
JOURNAL Nature 388 (6642), 539-547 (1997)
MEDLINE 97394467
REMARK Erratum:[[published erratum appears in Nature 1997 Sep 25;389(6649):412]]
REFERENCE 2 (bases 1 to 17880)
AUTHORS Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G., Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R., Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C., Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 17880)
AUTHORS White,O.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers

```
     source          1. .17880
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            complement(310. .1350)
                     /gene="HP1189"
     CDS             complement(310. .1350)
                     /gene="HP1189"
                     /note="similar to GB:L08471 SP:Q04797 GB:Z22554 PID:142828
                     PID:296147 percent identity: 45.67; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartate-semialdehyde dehydrogenase (asd)"
                     /protein_id="AAD08235.1"
                     /db_xref="GI:2314350"
                     /translation="MKTYNVAIVGASGAVGQELIKGLENSFFPIKKFVPLASTRSAGK
                     KIKAFNKDYEILETTHEVFEREKIDIAFFSAGGSVSEEFATSASKTALVVDNTSFFRL
                     NKDVPLVVPEINAKEIFNAPLNIIANPNCSTIQMTQILNPLHLHFKIKSVIVSTYQAV
                     SGAGNKGIESLKNELKTALECLEKDPTIDLNQVLQAGAFAYPIAFNAIAHIDTFKENG
                     YTKEELKMLHETHKIMGVDFPISATCVRVPVLRSHSESLSIAFEKEFDLKEVYEVLKN
                     APSVAVCDDPSHNLYPTPLKASHTDSVFIGRLRKDLFDKKTLHGFCVADQLRVGAATN
                     ALKIALHYIKNA"
     gene            complement(1337. .2665)
                     /gene="HP1190"
     CDS             complement(1337. .2665)
                     /gene="HP1190"
                     /note="similar to GB:Z11518 GB:X05345 GB:M96646 SP:P12081
                     PID:32460 percent identity: 32.39; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="histidyl-tRNA synthetase (hisS)"
                     /protein_id="AAD08236.1"
                     /db_xref="GI:2314351"
                     /translation="MITPKVLSGFKDRLPKDAIQKAQLLAKVSVVFQSFGFVPIETPH
                     LEYAQTLLPDASSDIQKEIYRFKDHGDRDVALRFDLTVPLARFVSLHHQTLGMPFKRY
                     AIGNVFRGERAQKGRYREFTQCDFDFIGSESLVCDAEIIQVIVASLKALDLEDFCVSI
                     NHRKILNGICEYFGISQVNEALRIVDKLEKIGLNGVEEELKKECGLNSNTIKELLELI
                     QIKQNDLSHAEFFEKIAYLKDYNENLKKGIQDLERLYQLLGDLQISQNLYKIDFSIAR
                     GLGYYTGIVYETTLNEMKSLGSVCSGGRYDHLTKNFSKENLQGVGASIGIDRLIVALS
                     EMQLLDERSTQAKVLIACMHEEYFSYANRLAESLRQSGIFSEVYPEAQKIKKPFSYAN
                     HKGHEFVAVIGEEEEFKSETLSLKNMHSGMQLNCLSFLKALEIIGENDEDL"
     gene            2727. .3776
                     /gene="HP1191"
     CDS             2727. .3776
                     /gene="HP1191"
                     /note="similar to GB:L42023 SP:P45042 PID:1006405
                     PID:1221223 PID:1205353 percent identity: 33.22;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ADP-heptose-lps heptosyltransferase II (rfaF)"
                     /protein_id="AAD08237.1"
                     /db_xref="GI:2314352"
                     /translation="MSVNAPKRMRILLRLPNWLGDGVMASSLFYTLKHHYPNAHFILV
                     GPTITCELFKKDEKIEAVFIDNTKKSFFRLLAIHKLAQKIGRCDIAITLNNHFYSAFL
                     LYATKTPVRIGFAQFFRSLFLSHAIAPAPKEYHQVEKYCFLFSQFLEKELDQKSVLPL
                     KLAFNLPTHTPNTPKKIGFNPSASYGSAKRWPASYYAEVSAVLLEKGHEIYFFGAKED
                     AIVSEEILKLIKGSLKNPSLFHNAYNLCGKTSIEELIERIAVLDLFITNDSGPMHVAA
                     SMQTPLIALFGPTDEKETRPYKAQKTIVLNHHLSCAPCKKRVCPLKNAKNHLCMKSIT
                     PLEVLEAAHTLLEEP"
     gene            complement(3967. .4248)
                     /gene="HP1192"
     CDS             complement(3967. .4248)
                     /gene="HP1192"
                     /note="similar to GP:1419559 percent identity: 72.53;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="secreted protein involved in flagellar motility"
                     /protein_id="AAD08238.1"
                     /db_xref="GI:2314353"
                     /translation="MKKQILTGVLLSVLAVSSAYAHKDKKDAKKPKFSTELVVAQNDK
                     KDAKKPKFSTELVVAQNDKKDAKKPKFSTELVVAQNDKKDAKKPKNSVV"
     gene            4466. .5455
                     /gene="HP1193"
     CDS             4466. .5455
                     /gene="HP1193"
                     /note="similar to SP:P40691 percent identity: 43.66;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aldo-keto reductase, putative"
                     /protein_id="AAD08243.1"
                     /db_xref="GI:2314358"
                     /translation="MQQRHLGPLKVGALALGCMGMTYGYGEVHDKKQMVKLIHKALEL
                     GINFFDTAEAYGEDNEKLLGEAIKPFKDKVVVASKFGIYYADPNDKYATMFLDSSPNR
                     IKSAIEGSLKRLKVECIDLYYQHRMDTNTPIGEVAEVMQLLIKEGKIKAWGMSEAGLS
                     SIQKAHQICPLSALQSEYSLWWREPEKEILGFLEKEKIGFVAFSPLGKGFLGAKFEKN
                     ATFASEDFRSVSPRFNQENLAKNYALVELIQDHAHAKGVTPAQLALSWILHTQKIIVP
                     LFGTTKESRLIENIGALQVSWSQKELEIFQKELTAIKIEGARYPERINEMVNQ"
     gene            complement(5626. .5712)
                     /gene="HP1194"
     CDS             complement(5626. .5712)
                     /gene="HP1194"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP1194"
                     /protein_id="AAD08244.1"
                     /db_xref="GI:2314359"
                     /translation="MLFKNPLIHRPPYKKFRLTNDKLFCIQS"
     gene            complement(5999. .8077)
                     /gene="HP1195"
     CDS             complement(5999. .8077)
                     /gene="HP1195"
```

```
                     /note="similar to GB:X00415 SP:P02996 GB:J01689 GB:V00356
                     PID:147895 percent identity: 67.48; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="translation elongation factor EF-G (fusA)"
                     /protein_id="AAD08239.1"
                     /db_xref="GI:2314354"
                     /translation="MARKTPLNRIRNIGIAAHIDAGKTTTSERILFYTGVSHKIGEVH
                     DGAATMDWMEQEKERGITITSAATTCFWKDHQINLIDTPGHVDFTIEVERSMRVLDGA
                     VSVFCSVGGVQPQSETVWRQANKYGVPRIVFVNKMDRIGANFYNVENQIKLRLKANPV
                     PINIPIGAEDTFIGVIDLVQMKAIVWNNETMGAKYDVEEIPSDLLEKAKEYREKLVEA
                     VAEQDEALMEKYLGGEELSIEEIKKGIKAGCLNMSLVPMLCGSSFKNKGVQTLLDAVI
                     DYLPAPTEVVDIKGIDPKTEEEVEFVKSSDDGEFAGLAFKIMTDPFVGQLTFVRVYRGK
                     LESGSYVYNSTKDKKERVGRLLKMHSNKREDIKEVYAGEICAFVGLKDTLTGDTLCDE
                     KNAVVLERMEFPEPVIHIAVEPKTKADQEKMGVALGKLAEEDPSFRVMTQEETGQTLI
                     GGMGELHLEIIVDRLKREFKVEAEIGQPQVAFRETIRSSVSKEHKYAKQSGGRGQYGH
                     VFIKLEPKEPGSGYEFVNEISGGVIPKEYIPAVDKGIQEAMQNGVLAGYPVVDFKVTL
                     YDGSYHDVDSSEMAFKIAGSMAFKEASRAANPVLLEPMMKVEVEVPEEYMGDVIGDLN
                     RRRGQINSMDDRLGLKIVNAFVPLVEMFGYSTDLRSATQGRGTYSMEFDHYGEVPSNI
                     AKEIVEKRKG"
     gene            complement(8089. .8556)
                     /gene="HP1196"
     CDS             complement(8089. .8556)
                     /gene="HP1196"
                     /note="similar to SP:P22744 percent identity: 61.44;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein S7 (rps7)"
                     /protein_id="AAD08240.1"
                     /db_xref="GI:2314355"
                     /translation="MRRRKAPVREVLGDPVYGNKVVTKFINKMMFDGKKSVAEKIIYK
                     AFNKIEEKSGEKGIEVFEKALERVRPLVEVRSRRVGGATYQVPVEVRASRQQSLSIRW
                     ILEATRKRNERMMVDRLANELMDAASDKGAAFKKKEDVHKMAEANKAFAHYRW"
     gene            complement(8572. .8979)
                     /gene="HP1197"
     CDS             complement(8572. .8979)
                     /gene="HP1197"
                     /note="similar to SP:P18662 percent identity: 79.03;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein S12 (rps12)"
                     /protein_id="AAD08241.1"
                     /db_xref="GI:2314356"
                     /translation="MPTINQLIRKERKKVVKKTKSPALVECPQRRGVCTRVYTTTPKK
                     PNSALRKVAKVRLTSKFEVISYIPGEGHNLQEHSIVLVRGGRVKDLPGVKYHIVRGAL

  Query Match             93.8%;  Score 15;  DB 1;  Length 17880;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaaccccc 16
           ||||||||||||||||
Db   16078 CTTTCACCAACCCCC 16092


RESULT  10
AC107593
LOCUS       AC107593               26236 bp    DNA     linear   HTG 23-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-186O5, *** SEQUENCING IN PROGRESS
            ***, 17 unordered pieces.
ACCESSION   AC107593
VERSION     AC107593.1  GI:18266586
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 26236)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 26236)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOVS
                Center clone name: CH230-186O5
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 98% of reads
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 21752 bases at least Q40
                Consensus quality: 23932 bases at least Q30
                Consensus quality: 25882 bases at least Q20
                Estimated insert size: 18630; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            * . (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1449: contig of 1449 bp in length
            *     1450     1549: gap of unknown length
            *     1550     4116: contig of 2567 bp in length
            *     4117     4216: gap of unknown length
            *     4217     6167: contig of 1951 bp in length
            *     6168     6267: gap of unknown length
            *     6268     7844: contig of 1577 bp in length
            *     7845     7944: gap of unknown length
            *     7945     9386: contig of 1442 bp in length
            *     9387     9486: gap of unknown length
            *     9487    11068: contig of 1582 bp in length
            *    11069    11168: gap of unknown length
            *    11169    12400: contig of 1232 bp in length
            *    12401    12500: gap of unknown length
            *    12501    13625: contig of 1125 bp in length
            *    13626    13725: gap of unknown length
            *    13726    15211: contig of 1486 bp in length
            *    15212    15311: gap of unknown length
            *    15312    16498: contig of 1187 bp in length
            *    16499    16598: gap of unknown length
            *    16599    17661: contig of 1063 bp in length
            *    17662    17761: gap of unknown length
            *    17762    19412: contig of 1651 bp in length
            *    19413    19512: gap of unknown length
            *    19513    20847: contig of 1335 bp in length
            *    20848    20947: gap of unknown length
            *    20948    22030: contig of 1083 bp in length
            *    22031    22130: gap of unknown length
            *    22131    23435: contig of 1305 bp in length
            *    23436    23535: gap of unknown length
            *    23536    24609: contig of 1074 bp in length
            *    24610    24709: gap of unknown length
            *    24710    26236: contig of 1527 bp in length.
FEATURES             Location/Qualifiers
     source          1. .26236
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-186O5"
BASE COUNT     7879 a   4887 c   4901 g   6949 t   1620 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 2;  Length 26236;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ctttcaccaacccccc 16
          ||||||||||||||||
Db   6101 CTTTCACCAACCCCC 6115


RESULT  11
AC005942
LOCUS       AC005942               26458 bp    DNA     linear   PRI 06-JUN-2000
DEFINITION  Homo sapiens chromosome Y, clone hCIT.298_B_15, complete sequence.
ACCESSION   AC005942
VERSION     AC005942.2  GI:4263846
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 26458)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Page,D. and Lander,E.
  TITLE     Homo sapiens chromosome Y, clone hCIT.298_B_15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 26458)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 26458)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 26458)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Skaletsky,H., Stange-Thomann,N., Stojanovic,N., Stone,C.,
            Subramanian,A., Sun,C., Tesfaye,S., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D.,
            Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 26458)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Skaletsky,H., Stange-Thomann,N., Stojanovic,N., Stone,C.,
            Subramanian,A., Sun,C., Tesfaye,S., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D.,
            Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 24, 1999 this sequence version replaced gi:4225939.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .26458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Y"
                     /map="Y"
                     /clone="hCIT.298_B_15"
                     /clone_lib="CITC Human BAC"
     repeat_region   complement(1. .123)
                     /rpt_family="AluY"
     repeat_region   complement(161. .437)
                     /rpt_family="LTR9"
     repeat_region   complement(415. .493)
                     /rpt_family="LTR9"
     repeat_region   complement(494. .736)
                     /rpt_family="AluSp"
     repeat_region   complement(737. .791)
                     /rpt_family="LTR9"
     unsure          complement(780. .821)
                     /note="Single-stranded coverage."
     repeat_region   complement(792. .1106)
                     /rpt_family="AluSq"
     repeat_region   complement(1107. .1392)
                     /rpt_family="LTR9"
     repeat_region   1396. .1784
                     /rpt_family="MER4B-internal"
     repeat_region   1785. .2001
                     /rpt_family="MER30"
     repeat_region   2002. .2098
                     /rpt_family="MER4B-internal"
     repeat_region   complement(2099. .2400)
                     /rpt_family="AluSx"
     repeat_region   2401. .2732
                     /rpt_family="MER4B-internal"
     repeat_region   complement(2734. .2921)
                     /rpt_family="THE1B"
     repeat_region   2898. .3456
                     /rpt_family="MER4A"
     repeat_region   complement(3480. .3538)
                     /rpt_family="MER72"
     repeat_region   3539. .3702
                     /rpt_family="MER4B-internal"
     repeat_region   complement(3715. .3974)
                     /rpt_family="MER4A"
     repeat_region   complement(4045. .4270)
                     /rpt_family="MER4A"
     repeat_region   4271. .4297
                     /rpt_family="(TTTG)n"
     repeat_region   complement(4298. .4393)
                     /rpt_family="MER4A"
     repeat_region   4413. .4506
                     /rpt_family="MER4B-internal"
     repeat_region   complement(4523. .4738)
                     /rpt_family="AluSq"
     repeat_region   4739. .4907
                     /rpt_family="MER4B-internal"
     repeat_region   complement(4908. .4926)
                     /rpt_family="MER50"
     repeat_region   4927. .5229
                     /rpt_family="AluSq"
     repeat_region   complement(5230. .5789)
                     /rpt_family="MER50"
     repeat_region   5792. .6033
                     /rpt_family="MER4D"
     repeat_region   6085. .6454
                     /rpt_family="MER4B-internal"
     repeat_region   6455. .6625
                     /rpt_family="MER4-internal"
     repeat_region   6631. .6844
                     /rpt_family="MER4-internal"
     repeat_region   complement(6845. .7155)
                     /rpt_family="AluSg"
     repeat_region   7156. .7355
                     /rpt_family="MER4-internal"
     repeat_region   7475. .7577
                     /rpt_family="MER4-internal"
     repeat_region   7798. .7890
                     /rpt_family="(TA)n"
     repeat_region   complement(8256. .8811)
                     /rpt_family="MER61E"
     repeat_region   9152. .9425
                     /rpt_family="AluJo"
     repeat_region   9509. .9568
                     /rpt_family="AT_rich"
     repeat_region   9848. .10031
                     /rpt_family="(TATATG)n"
     repeat_region   10219. .10440
                     /rpt_family="(TATATG)n"
     repeat_region   complement(10441. .10533)
                     /rpt_family="AluSg/x"
     repeat_region   11092. .11375
                     /rpt_family="AluSg"
     repeat_region   11376. .11398
                     /rpt_family="(CAAA)n"
     repeat_region   11425. .11451
                     /rpt_family="AT_rich"
     repeat_region   11961. .12074
                     /rpt_family="AluJo/FRAM"
     repeat_region   12180. .12206
                     /rpt_family="AT_rich"
     repeat_region   12320. .12403
                     /rpt_family="(TATG)n"
     repeat_region   complement(12784. .12954)
                     /rpt_family="AluSg/x"
     repeat_region   complement(13050. .13226)

  Query Match             93.8%;  Score 15;  DB 9;  Length 26458;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           ||||||||||||||||
Db   21811 CCTTTCACCAACCCC 21825


RESULT  12
HS212J10
LOCUS       HS212J10               30208 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from clone 212J10 on chromosome Xq25-26.3.
            Contains ESTs and a GSS, complete sequence.
ACCESSION   AL034401
VERSION     AL034401.1  GI:4071037
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 30208)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 29, 1998 this sequence version replaced gi:4007136.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            IMPORTANT: This sequence is not the entire insert of clone 212J10.
```

```
          It may be shorter because we only sequence overlapping sections
          once, or longer because we arrange for a small overlap between
          neighbouring submissions.
          The true left end of clone 448E20 (Z97196) is at 30109 in this
          sequence. The true right end of clone 417G15 (AL009174) is at 100
          in this sequence.
          This sequence has been finished according to sequence map criteria
          as follows. An attempt is made to resolve all sequencing problems,
          such as compressions and repeats, but not necessarily within known
          annotated human repeat sequence elements (e.g. Alu). Where the
          sequence is ambiguous, there is an annotation using the 'unsure'
          feature key.
          This sequence was generated from part of bacterial clone contigs of
          human chromosome X, constructed by the Sanger Centre Chromosome X
          Mapping Group. Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          212J10 is from the library RPCI1 constructed at the Roswell Park
          Cancer Institute by the group of Pieter de Jong. For further
          details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .30208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q25-26.3"
                     /clone="RP1-212J10"
                     /clone_lib="RPCI-1"
     repeat_region   1166. .1275
                     /note="L2 repeat: matches 2636. .2738 of consensus"
     repeat_region   1307. .1423
                     /note="MIR repeat: matches 19. .134 of consensus"
     repeat_region   1586. .1620
                     /note="L2 repeat: matches 2515. .2551 of consensus"
     repeat_region   1621. .1897
                     /note="AluSx repeat: matches 32. .312 of consensus"
     repeat_region   1898. .2108
                     /note="L2 repeat: matches 2551. .2745 of consensus"
     prim_transcript complement(2474. .2892)
                     /note="match: ESTs R94487 AI022438"
     misc_feature    3566. .4006
                     /note="match: GSS B95157"
     repeat_region   4688. .4806
                     /note="L2 repeat: matches 2575. .2698 of consensus"
     repeat_region   5684. .5993
                     /note="AluY repeat: matches 2. .311 of consensus"
     repeat_region   6908. .7088
                     /note="L2 repeat: matches 2495. .2679 of consensus"
     repeat_region   7298. .7608
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   8688. .8859
                     /note="MER45 repeat: matches 1. .173 of consensus"
     repeat_region   8860. .9149
                     /note="AluJb repeat: matches 1. .290 of consensus"
     repeat_region   9245. .9323
                     /note="L2 repeat: matches 2584. .2676 of consensus"
     repeat_region   12097. .12126
                     /note="10 copies 3 mer aac 100% conserved"
     repeat_region   12662. .12972
                     /note="AluSq repeat: matches 3. .309 of consensus"
     repeat_region   13719. .13770
                     /note="26 copies 2 mer ca 87% conserved"
     repeat_region   14015. .14135
                     /note="MER5B repeat: matches 12. .129 of consensus"
     repeat_region   14161. .14346
                     /note="L2 repeat: matches 1629. .1839 of consensus"
     repeat_region   15319. .15553
                     /note="MIR repeat: matches 20. .261 of consensus"
     repeat_region   16380. .16697
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   16805. .17219
                     /note="MSTB repeat: matches 1. .426 of consensus"
     repeat_region   17435. .17491
                     /note="MIR repeat: matches 201. .255 of consensus"
     repeat_region   17550. .17735
                     /note="MER5A repeat: matches 3. .189 of consensus"
     repeat_region   18324. .18488
                     /note="MIR repeat: matches 2. .218 of consensus"
     repeat_region   19212. .19275
                     /note="2 copies 32 mer 92% conserved"
     repeat_region   20376. .20588
                     /note="AluY repeat: matches 3. .215 of consensus"
     repeat_region   20589. .20630
                     /note="Alu repeat: matches 268. .307 of consensus"
     repeat_region   20774. .20914
                     /note="MER94 repeat: matches 1. .134 of consensus"
     repeat_region   21170. .21251
                     /note="41 copies 2 mer tt 76% conserved"
     repeat_region   24307. .24617
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   25794. .26104
                     /note="AluJo repeat: matches 1. .307 of consensus"
     repeat_region   26242. .26339
                     /note="MIR repeat: matches 164. .261 of consensus"
     repeat_region   26347. .26505
                     /note="L2 repeat: matches 1450. .1604 of consensus"
     repeat_region   28752. .29001
                     /note="L1PA3 repeat: matches 5894. .6146 of consensus"
     repeat_region   29377. .29494
                     /note="MIR repeat: matches 8. .124 of consensus"
     repeat_region   29510. .29698
                     /note="MIR repeat: matches 12. .212 of consensus"
BASE COUNT     9229 a   6309 c   5771 g   8899 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 9;  Length 30208;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           |||||||||||||||
Db    1968 CCTTTCACCAACCCC 1982


RESULT  13
AC099662
LOCUS       AC099662               46363 bp    DNA     linear   HTG 17-NOV-2001
DEFINITION  Rattus norvegicus clone CH230-1D24, *** SEQUENCING IN PROGRESS ***,
            40 unordered pieces.
ACCESSION   AC099662
VERSION     AC099662.1  GI:16973712
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 46363)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
```

```
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46363)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUFM
                Center clone name: CH230-1D24
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 99% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 19414 bases at least Q40
                Consensus quality: 22250 bases at least Q30
                Consensus quality: 24040 bases at least Q20
                Estimated insert size: 7952; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 40 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      897: contig of 897 bp in length
            *      898      997: gap of unknown length
            *      998     1915: contig of 918 bp in length
            *     1916     2015: gap of unknown length
            *     2016     2961: contig of 946 bp in length
            *     2962     3061: gap of unknown length
            *     3062     3970: contig of 909 bp in length
            *     3971     4070: gap of unknown length
            *     4071     4960: contig of 890 bp in length
            *     4961     5060: gap of unknown length
            *     5061     5989: contig of 929 bp in length
            *     5990     6089: gap of unknown length
            *     6090     7033: contig of 944 bp in length
            *     7034     7133: gap of unknown length
            *     7134     8079: contig of 946 bp in length
            *     8080     8179: gap of unknown length
            *     8180     9076: contig of 897 bp in length
            *     9077     9176: gap of unknown length
            *     9177    10077: contig of 901 bp in length
            *    10078    10177: gap of unknown length
            *    10178    11122: contig of 945 bp in length
            *    11123    11222: gap of unknown length
            *    11223    12158: contig of 936 bp in length
            *    12159    12258: gap of unknown length
            *    12259    13180: contig of 922 bp in length
            *    13181    13280: gap of unknown length
            *    13281    14190: contig of 910 bp in length
            *    14191    14290: gap of unknown length
            *    14291    16032: contig of 1742 bp in length
            *    16033    16132: gap of unknown length
            *    16133    17561: contig of 1429 bp in length
            *    17562    17661: gap of unknown length
            *    17662    17927: contig of 266 bp in length
            *    17928    18027: gap of unknown length
            *    18028    19429: contig of 1402 bp in length
            *    19430    19529: gap of unknown length
            *    19530    20481: contig of 952 bp in length
            *    20482    20581: gap of unknown length
            *    20582    21432: contig of 851 bp in length
            *    21433    21532: gap of unknown length
            *    21533    22815: contig of 1283 bp in length
            *    22816    22915: gap of unknown length
            *    22916    24228: contig of 1313 bp in length
            *    24229    24328: gap of unknown length
            *    24329    25333: contig of 1005 bp in length
            *    25334    25433: gap of unknown length
            *    25434    26350: contig of 917 bp in length
            *    26351    26450: gap of unknown length
            *    26451    27341: contig of 891 bp in length
            *    27342    27441: gap of unknown length
            *    27442    28606: contig of 1165 bp in length
            *    28607    28706: gap of unknown length
            *    28707    29621: contig of 915 bp in length
            *    29622    29721: gap of unknown length
            *    29722    30581: contig of 860 bp in length
            *    30582    30681: gap of unknown length
            *    30682    32084: contig of 1403 bp in length
            *    32085    32184: gap of unknown length
            *    32185    33096: contig of 912 bp in length
            *    33097    33196: gap of unknown length
            *    33197    34478: contig of 1282 bp in length
            *    34479    34578: gap of unknown length
            *    34579    36167: contig of 1589 bp in length
            *    36168    36267: gap of unknown length
            *    36268    37380: contig of 1113 bp in length
            *    37381    37480: gap of unknown length
            *    37481    38539: contig of 1059 bp in length
            *    38540    38639: gap of unknown length
            *    38640    40235: contig of 1596 bp in length
            *    40236    40335: gap of unknown length
            *    40336    41516: contig of 1181 bp in length
            *    41517    41616: gap of unknown length
            *    41617    42512: contig of 896 bp in length
            *    42513    42612: gap of unknown length
            *    42613    43771: contig of 1159 bp in length
            *    43772    43871: gap of unknown length
            *    43872    45209: contig of 1338 bp in length
            *    45210    45309: gap of unknown length
            *    45310    46363: contig of 1054 bp in length.
FEATURES             Location/Qualifiers
     source          1. .46363
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-1D24"
BASE COUNT    11211 a  11090 c   8068 g  11580 t   4414 others
ORIGIN
```

```
  Query Match             93.8%;  Score 15;  DB 2;  Length 46363;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           |||||||||||||||
Db   31978 CCTTTCACCAACCCC 31992


RESULT  14
AC110265
LOCUS       AC110265               62490 bp    DNA     linear   HTG 10-FEB-2002
DEFINITION  Mus musculus clone RP23-317B21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC110265
VERSION     AC110265.1  GI:18643635
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 62490)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-317B21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62490)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22627
                Center clone name: 317_B_21
            --------------.
            * NOTE: This record contains 79 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
*        1      650: contig of 650 bp in length
*      651 750: gap of      100 bp
*      751     1431: contig of 681 bp in length
*     1432 1531: gap of      100 bp
*     1532     2214: contig of 683 bp in length
*     2215 2314: gap of      100 bp
*     2315     3019: contig of 705 bp in length
*     3020 3119: gap of      100 bp
*     3120     3835: contig of 716 bp in length
*     3836 3935: gap of      100 bp
*     3936     4618: contig of 683 bp in length
*     4619 4718: gap of      100 bp
*     4719     5387: contig of 669 bp in length
*     5388 5487: gap of      100 bp
*     5488     6168: contig of 681 bp in length
*     6169 6268: gap of      100 bp
*     6269     6954: contig of 686 bp in length
*     6955 7054: gap of      100 bp
*     7055     7751: contig of 697 bp in length
*     7752 7851: gap of      100 bp
*     7852     8544: contig of 693 bp in length
*     8545 8644: gap of      100 bp
*     8645     9358: contig of 714 bp in length
*     9359 9458: gap of      100 bp
*     9459    10170: contig of 712 bp in length
*    10171 10270: gap of      100 bp
*    10271    10976: contig of 706 bp in length
*    10977 11076: gap of      100 bp
*    11077    11767: contig of 691 bp in length
*    11768 11867: gap of      100 bp
*    11868    12537: contig of 670 bp in length
*    12538 12637: gap of      100 bp
*    12638    13332: contig of 695 bp in length
*    13333 13432: gap of      100 bp
*    13433    14118: contig of 686 bp in length
*    14119 14218: gap of      100 bp
*    14219    14882: contig of 664 bp in length
*    14883 14982: gap of      100 bp
*    14983    15667: contig of 685 bp in length
*    15668 15767: gap of      100 bp
*    15768    16459: contig of 692 bp in length
*    16460 16559: gap of      100 bp
*    16560    17243: contig of 684 bp in length
*    17244 17343: gap of      100 bp
*    17344    18048: contig of 705 bp in length
*    18049 18148: gap of      100 bp
*    18149    18855: contig of 707 bp in length
*    18856 18955: gap of      100 bp
*    18956    19659: contig of 704 bp in length
*    19660 19759: gap of      100 bp
*    19760    20441: contig of 682 bp in length
*    20442 20541: gap of      100 bp
*    20542    21225: contig of 684 bp in length
*    21226 21325: gap of      100 bp
*    21326    22001: contig of 676 bp in length
*    22002 22101: gap of      100 bp
*    22102    22784: contig of 683 bp in length
*    22785 22884: gap of      100 bp
*    22885    23576: contig of 692 bp in length
*    23577 23676: gap of      100 bp
*    23677    24390: contig of 714 bp in length
*    24391 24490: gap of      100 bp
*    24491    25191: contig of 701 bp in length
*    25192 25291: gap of      100 bp
*    25292    26000: contig of 709 bp in length
*    26001 26100: gap of      100 bp
*    26101    26799: contig of 699 bp in length
*    26800 26899: gap of      100 bp
*    26900    27612: contig of 713 bp in length
*    27613 27712: gap of      100 bp
*    27713    28413: contig of 701 bp in length
*    28414 28513: gap of      100 bp
*    28514    29227: contig of 714 bp in length
```

```
          *    29228 29327: gap of      100 bp
          *    29328    30008: contig of 681 bp in length
          *    30009 30108: gap of      100 bp
          *    30109    30769: contig of 661 bp in length
          *    30770 30869: gap of      100 bp
          *    30870    31564: contig of 695 bp in length
          *    31565 31664: gap of      100 bp
          *    31665    32350: contig of 686 bp in length
          *    32351 32450: gap of      100 bp
          *    32451    33143: contig of 693 bp in length
          *    33144 33243: gap of      100 bp
          *    33244    33953: contig of 710 bp in length
          *    33954 34053: gap of      100 bp
          *    34054    34750: contig of 697 bp in length
          *    34751 34850: gap of      100 bp
          *    34851    35562: contig of 712 bp in length
          *    35563 35662: gap of      100 bp
          *    35663    36361: contig of 699 bp in length
          *    36362 36461: gap of      100 bp
          *    36462    37150: contig of 689 bp in length
          *    37151 37250: gap of      100 bp
          *    37251    37931: contig of 681 bp in length
          *    37932 38031: gap of      100 bp
          *    38032    38722: contig of 691 bp in length
          *    38723 38822: gap of      100 bp
          *    38823    39522: contig of 700 bp in length
          *    39523 39622: gap of      100 bp
          *    39623    40322: contig of 700 bp in length
          *    40323 40422: gap of      100 bp
          *    40423    41118: contig of 696 bp in length
          *    41119 41218: gap of      100 bp
          *    41219    41924: contig of 706 bp in length
          *    41925 42024: gap of      100 bp
          *    42025    42710: contig of 686 bp in length
          *    42711 42810: gap of      100 bp
          *    42811    43503: contig of 693 bp in length
          *    43504 43603: gap of      100 bp
          *    43604    44289: contig of 686 bp in length
          *    44290 44389: gap of      100 bp
          *    44390    45073: contig of 684 bp in length
          *    45074 45173: gap of      100 bp
          *    45174    45782: contig of 609 bp in length
          *    45783 45882: gap of      100 bp
          *    45883    46576: contig of 694 bp in length
          *    46577 46676: gap of      100 bp
          *    46677    47363: contig of 687 bp in length
          *    47364 47463: gap of      100 bp
          *    47464    48161: contig of 698 bp in length
          *    48162 48261: gap of      100 bp
          *    48262    48976: contig of 715 bp in length
          *    48977 49076: gap of      100 bp
          *    49077    49778: contig of 702 bp in length
          *    49779 49878: gap of      100 bp
          *    49879    50597: contig of 719 bp in length
          *    50598 50697: gap of      100 bp
          *    50698    51414: contig of 717 bp in length
          *    51415 51514: gap of      100 bp
          *    51515    52214: contig of 700 bp in length
          *    52215 52314: gap of      100 bp
          *    52315    53010: contig of 696 bp in length
          *    53011 53110: gap of      100 bp
          *    53111    53796: contig of 686 bp in length
          *    53797 53896: gap of      100 bp
          *    53897    54584: contig of 688 bp in length
          *    54585 54684: gap of      100 bp

  Query Match             93.8%;  Score 15;  DB 2;  Length 62490;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           |||||||||||||||
Db   48418 CCTTTCACCAACCCC 48432
```

```
RESULT  15
AC098003
LOCUS       AC098003               76122 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-103F13, *** SEQUENCING IN PROGRESS
            ***, 45 unordered pieces.
ACCESSION   AC098003
VERSION     AC098003.4  GI:17954427
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 76122)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76122)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17062261.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFWG
                Center clone name: CH230-103F13
            -------------- Summary Statistics
```

```
    Assembly program: Phrap; version 0.990329First call to
findPhrapList
    Consensus quality: 77971 bases at least Q40
    Consensus quality: 84068 bases at least Q30
    Consensus quality: 87671 bases at least Q20
    Estimated insert size: 56192; sum-of-contigs estimation
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     3003: contig of 3003 bp in length
*     3004     3103: gap of unknown length
*     3104     5847: contig of 2744 bp in length
*     5848     5947: gap of unknown length
*     5948     7577: contig of 1630 bp in length
*     7578     7677: gap of unknown length
*     7678    10093: contig of 2416 bp in length
*    10094    10193: gap of unknown length
*    10194    12839: contig of 2646 bp in length
*    12840    12939: gap of unknown length
*    12940    14522: contig of 1583 bp in length
*    14523    14622: gap of unknown length
*    14623    16376: contig of 1754 bp in length
*    16377    16476: gap of unknown length
*    16477    17626: contig of 1150 bp in length
*    17627    17726: gap of unknown length
*    17727    19448: contig of 1722 bp in length
*    19449    19548: gap of unknown length
*    19549    21139: contig of 1591 bp in length
*    21140    21239: gap of unknown length
*    21240    23330: contig of 2091 bp in length
*    23331    23430: gap of unknown length
*    23431    25797: contig of 2367 bp in length
*    25798    25897: gap of unknown length
*    25898    27622: contig of 1725 bp in length
*    27623    27722: gap of unknown length
*    27723    28928: contig of 1206 bp in length
*    28929    29028: gap of unknown length
*    29029    30615: contig of 1587 bp in length
*    30616    30715: gap of unknown length
*    30716    32381: contig of 1666 bp in length
*    32382    32481: gap of unknown length
*    32482    34314: contig of 1833 bp in length
*    34315    34414: gap of unknown length
*    34415    35724: contig of 1310 bp in length
*    35725    35824: gap of unknown length
*    35825    37981: contig of 2157 bp in length
*    37982    38081: gap of unknown length
*    38082    39099: contig of 1018 bp in length
*    39100    39199: gap of unknown length
*    39200    40816: contig of 1617 bp in length
*    40817    40916: gap of unknown length
*    40917    43194: contig of 2278 bp in length
*    43195    43294: gap of unknown length
*    43295    45001: contig of 1707 bp in length
*    45002    45101: gap of unknown length
*    45102    46852: contig of 1751 bp in length
*    46853    46952: gap of unknown length
*    46953    48710: contig of 1758 bp in length
*    48711    48810: gap of unknown length
*    48811    50063: contig of 1253 bp in length
*    50064    50163: gap of unknown length
*    50164    51383: contig of 1220 bp in length
*    51384    51483: gap of unknown length
*    51484    52716: contig of 1233 bp in length
*    52717    52816: gap of unknown length
*    52817    53858: contig of 1042 bp in length
*    53859    53958: gap of unknown length
*    53959    55844: contig of 1886 bp in length
*    55845    55944: gap of unknown length
*    55945    57056: contig of 1112 bp in length
*    57057    57156: gap of unknown length
*    57157    58388: contig of 1232 bp in length
*    58389    58488: gap of unknown length
*    58489    59718: contig of 1230 bp in length
*    59719    59818: gap of unknown length
*    59819    61042: contig of 1224 bp in length
*    61043    61142: gap of unknown length
*    61143    62324: contig of 1182 bp in length
*    62325    62424: gap of unknown length
*    62425    63845: contig of 1421 bp in length
*    63846    63945: gap of unknown length
*    63946    65202: contig of 1257 bp in length
*    65203    65302: gap of unknown length
*    65303    66761: contig of 1459 bp in length
*    66762    66861: gap of unknown length
*    66862    67864: contig of 1003 bp in length
*    67865    67964: gap of unknown length
*    67965    69347: contig of 1383 bp in length
*    69348    69447: gap of unknown length
*    69448    70673: contig of 1226 bp in length
*    70674    70773: gap of unknown length
*    70774    72161: contig of 1388 bp in length
*    72162    72261: gap of unknown length
*    72262    73645: contig of 1384 bp in length
*    73646    73745: gap of unknown length
*    73746    75004: contig of 1259 bp in length
*    75005    75104: gap of unknown length
*    75105    76122: contig of 1018 bp in length.
FEATURES             Location/Qualifiers
     source          1. .76122
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-103F13"
BASE COUNT    22090 a  13164 c  13749 g  22647 t   4472 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 2;  Length 76122;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaaccccc 16
           |||||||||||||||
Db   73808 CTTTCACCAACCCCC 73822



Search completed: July 29, 2002, 23:55:03
Job time: 7019 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 22:39:09 ; Search time 285.14 Seconds
(without alignments)
96.341 Million cell updates/sec

Title: US-09-530-663B-15
Perfect score: 16
Sequence: 1 cctttcaccaaccccc 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		16	100.0	1172	20	AAX58644	Aspen 4-coumarate
2		15	93.8	1440	6	AAN50528	Parasponia rhizobi
c 3		15	93.8	3084	19	AAX13959	H. pylori GHPO 741
c 4		15	93.8	8673	23	AAS53676	Helicobacter pylor
c 5		15	93.8	335913	22	AAI61371	Soybean 240017 reg
c 6		15	93.8	335913	22	AAI61372	Soybean 240017 reg
c 7		14.4	90.0	347	22	AAD05460	Human secreted pro
c 8		14.4	90.0	411	22	AAI83975	Human polynucleoti
c 9		14.4	90.0	457	22	ABA11306	Human nervous syst
c 10		14.4	90.0	606	23	AAS64618	DNA encoding novel
c 11		14.4	90.0	736	22	AAK70540	Human immune/haema
c 12		14.4	90.0	736	22	AAK70541	Human immune/haema
c 13		14.4	90.0	736	22	AAK70542	Human immune/haema
14		14.4	90.0	763	22	AAL15377	Human breast cance
c 15		14.4	90.0	807	22	AAH07529	Human cDNA clone (
c 16		14.4	90.0	896	21	AAZ50918	Human Protease and
c 17		14.4	90.0	1323	22	AAK81345	Human immune/haema
18		14.4	90.0	1514	19	AAV65632	Human heart O-fuco
19		14.4	90.0	1789	21	AAC51694	Zea mays DNA fragm
c 20		14.4	90.0	2093	22	AAH14829	Human cDNA sequenc
c 21		14.4	90.0	2159	24	AAS62693	cDNA sequence #480
22		14.4	90.0	2402	22	AAK86348	Human immune/haema
23		14.4	90.0	2823	22	AAF57033	DNA sequence of re
24		14.4	90.0	4306	23	AAS74164	DNA encoding novel
25		14.4	90.0	4453	22	AAK51977	Human polynucleoti
c 26		14.4	90.0	4454	22	AAK52961	Human polynucleoti
27		14.4	90.0	4850	22	AAK52494	Human polynucleoti
28		14.4	90.0	5009	19	AAV65634	First EcoR1 nucleo
29		14.4	90.0	5230	22	AAK51510	Human polynucleoti
c 30		14.4	90.0	6694	24	AAS61312	Human gene regulat
c 31		14.4	90.0	10713	24	ABL32740	Human immune syste
32		14.4	90.0	11284	19	AAV65633	Plasmid construct
c 33		14.4	90.0	11534	24	ABL32343	Human immune syste
c 34		14.4	90.0	12359	24	ABL34046	Human immune syste
35		14.4	90.0	28676	22	AAK80349	Human immune/haema
36		14.4	90.0	32222	22	AAK91316	Human digestive sy
37		14.4	90.0	32222	22	AAS32143	Human liver associ
c 38		14.4	90.0	35633	23	ABL09356	Drosophila melanog
c 39		14.4	90.0	349980	22	AAH41224	Pyrococcus abyssi
c 40		14.4	90.0	349980	22	AAH41225	Pyrococcus abyssi
c 41		14	87.5	384	14	AAQ60203	Human brain Expres
42		14	87.5	2811	23	AAS78129	DNA encoding novel
43		14	87.5	7780	21	AAA60630	HNRCR nucleotide s
c 44		14	87.5	16373	24	ABL32619	Human immune syste
c 45		14	87.5	17934	24	ABL33719	Human immune syste

ALIGNMENTS

RESULT 1
AAX58644
ID AAX58644 standard; cDNA; 1172 BP.
XX
AC AAX58644;
XX
DT 16-AUG-1999 (first entry)
XX
DE Aspen 4-coumarate coenzyme A ligase Pt4CL1 gene promoter.
XX
KW Aspen; 4-coumarate coenzyme A ligase; lignin; transgenic plant;
KW tree; conifer; forestry; Pt4CL1; promoter; xylem; ss.
XX
OS Populus tremuloides.
XX
FH Key Location/Qualifiers
FT misc_feature 888..893
FT /*tag= a
FT /note= "box A, cis-acting element"
FT misc_feature 914..924
FT /*tag= b
FT /note= "box L, cis-acting element"
FT misc_feature 1018..1033
FT /*tag= c
FT /note= "box P, cis-acting element"
XX
PN WO9924561-A2.
XX
PD 20-MAY-1999.
XX
PF 12-NOV-1998; 98WO-US24138.
XX

PR 12-NOV-1997; 97US-0969046.
XX
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Chiang VLC, Hu W, Tsai C;
XX
DR WPI; 1999-327394/27.
XX
PT Altering properties of plants by modulating 4-coumarate co-enzyme A
PT ligase
XX
PS Disclosure; Page 71; 73pp; English.
XX
CC This is the nucleotide sequence of the promoter region of the
CC quaking aspen (Populus tremuloides Michx.) Pt4CL1 gene (see
CC also AAX58642) that codes for 4-coumarate coenzyme A ligase 4CL1 (see
CC AAY06092). The promoter DNA was isolated from an aspen genomic
CC library by screening with Pt4CL1 cDNA. The promoter drives gene
CC expression exclusively in xylem tissue. It can be used to
CC manipulate gene expression, and hence to engineer traits of
CC interest, in the xylem tissue of target plants, e.g. to manipulate
CC lignin content or structure, or to enhance growth, cellulose
CC content or other value-added wood qualities. Plants with altered
CC contents of lignin and/or cellulose can be processed more
CC efficiently, e.g. for pulp production, with reduced costs and
CC pollution associated with lignin removal.
XX
SQ Sequence 1172 BP; 399 A; 225 C; 180 G; 368 T; 0 other;

Query Match 100.0%; Score 16; DB 20; Length 1172;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||||||
Db 1018 cctttcaccaaccccc 1033

RESULT 2
AAN50528
ID AAN50528 standard; DNA; 1440 BP.
XX
AC AAN50528;
XX
DT 09-DEC-1991 (first entry)
XX
DE Parasponia rhizobium ANU289 nifD gene.
XX
KW nifD gene; nitrogenase; enzyme; nitrogen-fixation;
KW strain improvement; plasmid pBR289nif-2; plasmid pRt329nif-2; ss.
XX
OS Parasponia rhizobium.
XX
FH Key Location/Qualifiers
FT CDS 175..1434
FT /*tag= a
XX
PN EP130047-A.
XX
PD 02-JAN-1985.
XX
PF 21-JUN-1984; 84EP-0304191.
XX
PR 22-JUN-1983; 83US-0506676.
PR 16-OCT-1987; 87US-0109868.
XX
PA (AGRI-) AGRIGENETICS RES AS.
XX
PI Shine J, Rolfe BG, Scott KF;
XX
DR WPI; 1985-007981/02.
DR P-PSDB; AAP50775.
XX
PT Bacterial strain contg. recombinant DNA fragment - esp. in
PT Rhizobium strains for improved nitrogen fixation
XX
PS Disclosure; Fig 5; 52pp; English.
XX
CC When this gene is expressed in a Rhizobium sp., the bacterium
CC produces useful products and plants may have their properties
CC improved, e.g. the rate, quality and efficiency of the nitrogen-
CC fixation process, especially in the root nodules of Rhizobium
CC strains. See also AAN50526-7 and AAP50773-4.
XX
SQ Sequence 1440 BP; 355 A; 509 C; 283 G; 293 T; 0 other;

Query Match 93.8%; Score 15; DB 6; Length 1440;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctttcaccaaccccc 16
|||||||||||||||
Db 294 ctttcaccaaccccc 308

RESULT 3
AAX13959/c
ID AAX13959 standard; DNA; 3084 BP.
XX
AC AAX13959;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 741 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 49..3030
FT /*tag= a
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR P-PSDB; AAW98240.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 200-204; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used

CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 3084 BP; 960 A; 589 C; 721 G; 814 T; 0 other;

Query Match 93.8%; Score 15; DB 19; Length 3084;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctttcaccaaccccc 16
||||||||||||||||
Db 1711 CTTTCACCAACCCCC 1697

RESULT 4
AAS53676/c
ID AAS53676 standard; DNA; 8673 BP.
XX
AC AAS53676;
XX
DT 13-FEB-2002 (first entry)
XX
DE Helicobacter pylori DNA for cellular proliferation protein #130.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Helicobacter pylori.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU35817.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7313; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 8673 BP; 2673 A; 1595 C; 2181 G; 2224 T; 0 other;

Query Match 93.8%; Score 15; DB 23; Length 8673;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctttcaccaaccccc 16
||||||||||||||||
Db 1663 CTTTCACCAACCCCC 1649

RESULT 5
AAI61371/c
ID AAI61371 standard; DNA; 335913 BP.
XX
AC AAI61371;
XX
DT 16-OCT-2001 (first entry)
XX
DE Soybean 240017 region G3, SEQ ID NO: 2.
XX
KW Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS Glycine max.
XX
PN WO200151627-A2.
XX
PD 19-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-US00552.
XX
PR 07-JAN-2000; 2000US-0174880.
XX
PA (MONS) MONSANTO CO.
XX
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
DR WPI; 2001-425872/45.
DR P-PSDB; AAM42214.
XX
PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
XX
PS Claim 2; Page 204-400; 1353pp; English.
XX
CC The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
CC provided in the specification.
XX
SQ Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 other;

Query Match 93.8%; Score 15; DB 22; Length 335913;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctttcaccaacccc 15
|||||||||||||||

```
Db 291134 CCTTTCACCAACCCC 291120


RESULT   6
AAI61372/c
ID   AAI61372 standard; DNA; 335913 BP.
XX
AC   AAI61372;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 240017 region G3, SEQ ID NO: 3.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US00552.
XX
PR   07-JAN-2000; 2000US-0174880.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
DR   P-PSDB; AAM42215.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs -
XX
PS   Claim 2; Page 400-595; 1353pp; English.
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome, which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule
CC   provided in the specification.
XX
SQ   Sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 other;


  Query Match             93.8%;  Score 15;  DB 22;  Length 335913;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           |||||||||||||||
Db 291134 CCTTTCACCAACCCC 291120


RESULT   7
AAD05460/c
ID   AAD05460 standard; cDNA; 347 BP.
XX
AC   AAD05460;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 22 cDNA clone HT4ES80, SEQ ID NO:82.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; chromosome 16; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..213
FT                   /*tag=  a
FT                   /product= "Human secreted protein"
FT                   /transl_except= (pos:58..60, aa:Xaa)
FT                   /transl_except= (pos:178..180, aa:Xaa)
FT                   /note= "Xaa corresponds to any of the naturally occurring
FT                   L-amino acids; CDS does not include start codon"
FT   sig_peptide     1..3
FT                   /*tag=  b
FT   mat_peptide     4..210
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200134623-A1.
XX
PD   17-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30037.
XX
PR   05-NOV-1999;   99US-0163577.
PR   30-JUN-2000; 2000US-0215137.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Moore PA;
XX
DR   WPI; 2001-316490/33.
DR   P-PSDB; AAE01617.
XX
PT   Nucleic acids encoding 29 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 1; Page 468; 535pp; English.
XX
CC   AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
CC   protein genes, and AAE01546-AAE01630 represent the proteins they encode.
CC   AAE01631-AAE01660 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   29 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
```

```
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 347 BP; 73 A; 87 C; 106 G; 76 T; 5 other;


  Query Match             90.0%;  Score 14.4;  DB 22;  Length 347;
  Best Local Similarity   93.8%;  Pred. No. 3.9e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
           |||||| |||||||||
Db      31 CCTTTCTCCAACCCCC 16


RESULT   8
AAI83975/c
ID   AAI83975 standard; cDNA; 411 BP.
XX
AC   AAI83975;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4035.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO04044.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 4035; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 411 BP; 149 A; 73 C; 116 G; 72 T; 1 other;


  Query Match             90.0%;  Score 14.4;  DB 22;  Length 411;
  Best Local Similarity   93.8%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
           |||||| |||||||||
Db     173 CCTTTCCCCAACCCCC 158


RESULT   9
ABA11306/c
ID   ABA11306 standard; cDNA; 457 BP.
XX
AC   ABA11306;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 313.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
```

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR P-PSDB; ABB14980.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 1; SEQ ID NO 313; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 457 BP; 141 A; 81 C; 86 G; 144 T; 5 other;

Query Match 90.0%; Score 14.4; DB 22; Length 457;
Best Local Similarity 93.8%; Pred. No. 4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||| |||
Db 457 CCTTTCACCAACACCC 442

```
RESULT  10
AAS64618/c
ID   AAS64618 standard; cDNA; 606 BP.
XX
AC   AAS64618;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #422.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00431.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 422; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 606 BP; 128 A; 188 C; 133 G; 157 T; 0 other;


  Query Match             90.0%;  Score 14.4;  DB 23;  Length 606;
  Best Local Similarity   93.8%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||| ||||
Db    532 CCTTTCACCAATCCCC 517


RESULT  11
```

```
AAK70540/c
ID   AAK70540 standard; DNA; 736 BP.
XX
AC   AAK70540;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25352.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
```

PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 25352; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 736 BP; 125 A; 218 C; 192 G; 201 T; 0 other;

Query Match 90.0%; Score 14.4; DB 22; Length 736;
Best Local Similarity 93.8%; Pred. No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||||||
Db 571 CCTTTCCCCAACCCCC 556

RESULT 12
AAK70541/c
ID AAK70541 standard; DNA; 736 BP.
XX
AC AAK70541;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25353.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and

PT metastasis -
XX
PS Disclosure; SEQ ID NO 25353; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 736 BP; 126 A; 218 C; 191 G; 201 T; 0 other;

Query Match 90.0%; Score 14.4; DB 22; Length 736;
Best Local Similarity 93.8%; Pred. No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||| ||||||||
Db 571 CCTTTCCCCAACCCCC 556

RESULT 13
AAK70542/c
ID AAK70542 standard; DNA; 736 BP.
XX
AC AAK70542;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25354.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

```
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 25354; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 736 BP; 126 A; 212 C; 193 G; 205 T; 0 other;


  Query Match             90.0%;  Score 14.4;  DB 22;  Length 736;
  Best Local Similarity   93.8%;  Pred. No. 4.1e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
           |||||| |||||||||
Db     571 CCTTTCCCCAACCCCC 556


RESULT  14
AAL15377
ID   AAL15377 standard; cDNA; 763 BP.
XX
AC   AAL15377;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 7834.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 1411; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 763 BP; 178 A; 183 C; 219 G; 153 T; 30 other;


  Query Match             90.0%;  Score 14.4;  DB 22;  Length 763;
  Best Local Similarity   93.8%;  Pred. No. 4.1e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
           |||||| |||||||||
Db       1 cctttccccaaccccc 16
```

RESULT 15
AAH07529/c
ID AAH07529 standard; cDNA; 807 BP.
XX
AC AAH07529;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4364.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 4364; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 807 BP; 240 A; 120 C; 167 G; 274 T; 6 other;

Query Match 90.0%; Score 14.4; DB 22; Length 807;
Best Local Similarity 93.8%; Pred. No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
|||||||| ||||||||
Db 555 CCTTTCATCAACCCCC 540

Search completed: July 30, 2002, 00:01:18
Job time: 4929 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 21:29:04 ; Search time 2542.47 Seconds
(without alignments)
84.938 Million cell updates/sec

Title: US-09-530-663B-15
Perfect score: 16
Sequence: 1 cctttcaccaaccccc 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16	100.0	218	9	BB430167	BB430167 BB430167
c	2	16	100.0	445	9	AI452374	AI452374 mq89g07.x
	3	16	100.0	564	12	AZ871533	AZ871533 2M0184N21
c	4	16	100.0	796	12	BH557406	BH557406 BOGJO56TR
c	5	16	100.0	837	10	BG771897	BG771897 602721709
c	6	16	100.0	1024	12	AZ210049	AZ210049 SP_0153_A
c	7	15	93.8	190	12	AZ299798	AZ299798 RPCI-23-4
c	8	15	93.8	246	12	AZ019845	AZ019845 RPCI-23-2
c	9	15	93.8	329	9	AA503034	AA503034 nh58b03.s
c	10	15	93.8	353	12	B82624	B82624 RPCI11-17C1
	11	15	93.8	372	10	BF766646	BF766646 IL2-CS004
	12	15	93.8	380	9	AV212743	AV212743 AV212743
	13	15	93.8	408	10	BM106270	BM106270 509959 MA
c	14	15	93.8	425	9	AI991910	AI991910 ws42c04.x
c	15	15	93.8	507	12	AQ880479	AQ880479 HS_5044_A
c	16	15	93.8	626	10	BE347738	BE347738 sp02g09.y
c	17	15	93.8	627	9	AV679827	AV679827 AV679827
c	18	15	93.8	630	10	BE556267	BE556267 sp99e11.y
c	19	15	93.8	636	10	BF129318	BF129318 601810938
	20	15	93.8	662	12	AQ794741	AQ794741 nbxb0053H
c	21	15	93.8	704	12	AG166290	AG166290 Pan trogl
c	22	15	93.8	840	10	BG395823	BG395823 602458604
	23	15	93.8	866	12	AQ271654	AQ271654 nbxb0026B
c	24	15	93.8	867	10	BI953359	BI953359 HVSMEm001
	25	15	93.8	901	10	BG167717	BG167717 602342919
	26	15	93.8	910	10	BE414064	BE414064 SCU005.H0
c	27	15	93.8	935	10	BF788352	BF788352 602114092
c	28	15	93.8	938	10	BE285548	BE285548 601095546
c	29	15	93.8	1045	10	BE966779	BE966779 601661280
	30	15	93.8	1079	10	BG338822	BG338822 602436605
	31	14.4	90.0	128	9	AV415808	AV415808 AV415808
c	32	14.4	90.0	159	9	BB593966	BB593966 BB593966
c	33	14.4	90.0	166	10	BG964367	BG964367 602831984
	34	14.4	90.0	167	9	AV313903	AV313903 AV313903
	35	14.4	90.0	179	9	AV029522	AV029522 AV029522
	36	14.4	90.0	185	10	BI202286	BI202286 t2h05fs.f
c	37	14.4	90.0	186	10	C88801	C88801 C88801 Mous
c	38	14.4	90.0	193	9	AV092502	AV092502 AV092502
c	39	14.4	90.0	197	10	BG981061	BG981061 MR3-CN014
c	40	14.4	90.0	200	9	AU180999	AU180999 AU180999
c	41	14.4	90.0	204	9	BB600337	BB600337 BB600337
	42	14.4	90.0	210	9	AV354827	AV354827 AV354827
	43	14.4	90.0	223	9	BB496202	BB496202 BB496202
	44	14.4	90.0	230	9	BB442511	BB442511 BB442511
	45	14.4	90.0	231	9	BB514048	BB514048 BB514048

ALIGNMENTS

RESULT 1
BB430167
LOCUS BB430167 218 bp mRNA linear EST 18-JUL-2000
DEFINITION BB430167 RIKEN full-length enriched, adult male hippocampus Mus musculus cDNA clone C630031O19 3' similar to X86406 R.norvegicus mRNA for brevican, GPI-anchored isoform, mRNA sequence.
ACCESSION BB430167
VERSION BB430167.1 GI:9269894
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 218)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .218
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C630031O19"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hippocampus"
                     /sex="male"
                     /tissue_type="hippocampus"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']"
BASE COUNT       52 a     73 c     31 g     62 t
ORIGIN
```

```
  Query Match           100.0%;  Score 16;  DB 9;  Length 218;
  Best Local Similarity  100.0%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db     49 CCTTTCACCAACCCCC 64
```

```
RESULT   2
AI452374/c
LOCUS       AI452374                 445 bp    mRNA    linear   EST 09-MAR-1999
DEFINITION  mq89g07.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:585948 3', mRNA sequence.
ACCESSION   AI452374
VERSION     AI452374.1  GI:4301476
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:585948"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       87 a     105 c     151 g     100 t       2 others
ORIGIN
```

```
  Query Match           100.0%;  Score 16;  DB 9;  Length 445;
  Best Local Similarity  100.0%;  Pred. No. 2.4e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db    394 CCTTTCACCAACCCCC 379
```

```
RESULT   3
AZ871533
LOCUS       AZ871533                 564 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0184N21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0184N21 F, DNA sequence.
ACCESSION   AZ871533
VERSION     AZ871533.1  GI:13077828
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0184  row: N  column: 21
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 564.
FEATURES             Location/Qualifiers
     source          1. .564
```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0184N21"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 143 a 123 c 103 g 195 t
ORIGIN

Query Match 100.0%; Score 16; DB 12; Length 564;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||||||
Db 165 CCTTTCACCAACCCCC 180

RESULT 4
BH557406/c
LOCUS BH557406 796 bp DNA linear GSS 14-DEC-2001
DEFINITION BOGJO56TR BOGJ Brassica oleracea genomic clone BOGJO56, DNA sequence.
ACCESSION BH557406
VERSION BH557406.1 GI:17809186
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 796)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGJO56TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .796
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGJO56"
/clone_lib="BOGJ"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 199 a 137 c 254 g 206 t
ORIGIN

Query Match 100.0%; Score 16; DB 12; Length 796;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||||||
Db 678 CCTTTCACCAACCCCC 663

RESULT 5
BG771897/c
LOCUS BG771897 837 bp mRNA linear EST 15-MAY-2001
DEFINITION 602721709F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838306 5', mRNA sequence.
ACCESSION BG771897
VERSION BG771897.1 GI:14082550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 837)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10772 row: b column: 03
High quality sequence stop: 837.
FEATURES Location/Qualifiers
source 1. .837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838306"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 226 a 179 c 249 g 183 t
ORIGIN

Query Match 100.0%; Score 16; DB 10; Length 837;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cctttcaccaaccccc 16
||||||||||||||||
Db 663 CCTTTCACCAACCCCC 648

RESULT 6

```
AZ210049/c
LOCUS       AZ210049                1024 bp    DNA     linear   GSS 31-AUG-2000
DEFINITION  SP_0153_A1_H11_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=153 Col=21 Row=O, DNA sequence.
ACCESSION   AZ210049
VERSION     AZ210049.1  GI:8424406
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 1024)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 153  row: O  column: 21
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 1024.
FEATURES             Location/Qualifiers
     source          1. .1024
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=153 Col=21 Row=O"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      301 a    187 c    212 g    324 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 12;  Length 1024;
  Best Local Similarity  100.0%;  Pred. No. 2.4e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||||||
Db   1020 CCTTTCACCAACCCCC 1005


RESULT   7
AZ299798/c
LOCUS       AZ299798                 190 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-439C15.TV RPCI-23 Mus musculus genomic clone RPCI-23-439C15
            , DNA sequence.
ACCESSION   AZ299798
VERSION     AZ299798.1  GI:9541583
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-439C15.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 439  row: C  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-439C15"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      35 a     35 c     36 g     84 t
ORIGIN


  Query Match           93.8%;  Score 15;  DB 12;  Length 190;
  Best Local Similarity  100.0%;  Pred. No. 6.4e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccc 15
          |||||||||||||||
Db    109 CCTTTCACCAACCCC 95


RESULT   8
AZ019845/c
LOCUS       AZ019845                 246 bp    DNA     linear   GSS 25-FEB-2000
DEFINITION  RPCI-23-294J24.TJ RPCI-23 Mus musculus genomic clone RPCI-23-294J24
            , DNA sequence.
ACCESSION   AZ019845
VERSION     AZ019845.1  GI:7095229
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 246)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
```

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 294  row: J  column: 24
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .246
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-23-294J24"
                      /clone_lib="RPCI-23"
                      /sex="Female"
                      /lab_host="DH10B"
                      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                      EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                      brain genomic DNA was isolated and  partially digested
                      with a combination of EcoRI and EcoRI Methylase.   Size
                      selected DNA was cloned into the pBACe3.6 vector at the
                      EcoRI  sites.  The ligation products were transformed into
                      DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       49 a     41 c     48 g    108 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 12;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          |||||||||||||||
Db    108 CCTTTCACCAACCCC 94


RESULT  9
AA503034/c
LOCUS       AA503034                329 bp    mRNA    linear   EST 19-AUG-1997
DEFINITION  nh58b03.s1 NCI_CGAP_Pr8 Homo sapiens cDNA clone IMAGE:956525
            similar to gb:D10667 MYOSIN HEAVY CHAIN, SMOOTH MUSCLE ISOFORM
            (HUMAN);contains element PTR7 repetitive element ;, mRNA sequence.
ACCESSION   AA503034
VERSION     AA503034.1  GI:2238001
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 329)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .329
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:956525"
                      /clone_lib="NCI_CGAP_Pr8"
                      /sex="male"
                      /tissue_type="prostate"
                      /lab_host="DH10B"
                      /note="Vector: pAMP10; mRNA made from invasive prostate
                      tumor, cDNA made by oligo-dT priming.   Non-directionally
                      cloned.  Size-selected on agarose gel, average insert
                      size 600 bp.  "
BASE COUNT       53 a     82 c     73 g    121 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 9;  Length 329;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ctttcaccaaccccc 16
          |||||||||||||||
Db    173 CTTTCACCAACCCCC 159


RESULT  10
B82624/c
LOCUS       B82624                  353 bp    DNA     linear   GSS 09-APR-1999
DEFINITION  RPCI11-17C17.TP RPCI-11 Homo sapiens genomic clone RPCI-11-17C17,
            DNA sequence.
ACCESSION   B82624
VERSION     B82624.1  GI:2869647
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-17C17.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Homo sapiens"
                      /db_xref="GDB:7506208"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-17C17"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      102 a     60 c     65 g    126 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 12;  Length 353;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          |||||||||||||||
```

```
Db    274 CCTTTCACCAACCCC 260


RESULT  11
BF766646
LOCUS       BF766646                372 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  IL2-CS0048-221100-249-D11 CS0048 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF766646
VERSION     BF766646.1  GI:12114546
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t2=IL2-CS0048-
            221100-249-D11&t3=2000-11-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 29
            High quality sequence stop: 372.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CS0048"
                     /dev_stage="Adult"
                     /note="Organ: colon_est; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       97 a    101 c     62 g    112 t
ORIGIN


  Query Match           93.8%;  Score 15;  DB 10;  Length 372;
  Best Local Similarity  100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          ||||||||||||||||
Db    353 CCTTTCACCAACCCC 367


RESULT  12
AV212743
LOCUS       AV212743                380 bp    mRNA    linear   EST 30-OCT-1999
DEFINITION  AV212743 RIKEN full-length enriched, ES cells Mus musculus cDNA
            clone 2410118E09 3' similar to D85818 Mouse mRNA for RNA polymerase
            II subuunit RPB14, mRNA sequence.
ACCESSION   AV212743
VERSION     AV212743.1  GI:6154362
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2410118E09"
                     /clone_lib="RIKEN full-length enriched, ES cells"
                     /cell_type="ES cells"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']."
BASE COUNT      105 a    118 c     71 g     86 t
ORIGIN


  Query Match           93.8%;  Score 15;  DB 9;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      2 ctttcaccaaccccc 16
          ||||||||||||||||
Db    131 CTTTCACCAACCCCC 145


RESULT  13
BM106270
LOCUS       BM106270                408 bp    mRNA    linear   EST 21-NOV-2001
DEFINITION  509959 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BM106270
VERSION     BM106270.1  GI:17037340
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 103  row: O  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .408
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 3BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from marrow, alveolar
                      macrophage, ovary, fetal semitendonosus muscle, and fetal
                      longissimus muscle."
BASE COUNT       94 a     94 c    102 g    118 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 10;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          ||||||||||||||||
Db    226 CCTTTCACCAACCCC 240


RESULT  14
AI991910/c
LOCUS       AI991910                425 bp    mRNA    linear   EST 08-MAR-2000
DEFINITION  ws42c04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2499846 3'
            similar to contains TAR1.t1 TAR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI991910
VERSION     AI991910.1  GI:5838815
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 822   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 416.
FEATURES              Location/Qualifiers
     source           1. .425
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2499846"
                      /clone_lib="NCI_CGAP_Brn25"
                      /tissue_type="anaplastic oligodendroglioma"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library is normalized, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT       83 a    151 c    125 g     63 t      3 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 9;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaacccc 15
          ||||||||||||||||
Db    358 CCTTTCACCAACCCC 344


RESULT  15
AQ880479/c
LOCUS       AQ880479                507 bp    DNA     linear   GSS 09-NOV-1999
DEFINITION  HS_5044_A2_G02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=8812 Col=4 Row=M, DNA sequence.
ACCESSION   AQ880479
VERSION     AQ880479.1  GI:6311946
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
```

```
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 8812  row: M  column: 4
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=8812 Col=4 Row=M"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      141 a    117 c    107 g    135 t      7 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 12;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 6.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaaccccc 16
           ||||||||||||||||
Db     385 CTTTCACCAACCCCC 371


Search completed: July 29, 2002, 23:22:40
Job time: 6816 sec
```

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 22:36:59 ; Search time 65.09 Seconds
(without alignments)
60.380 Million cell updates/sec

Title: US-09-530-663B-15
Perfect score: 16
Sequence: 1 cctttcaccaaccccc 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	14.4	90.0	1100	3	US-08-978-741-16	Sequence 16, Appl
2	14.4	90.0	1100	4	US-09-333-729A-16	Sequence 16, Appl
3	14.4	90.0	1300	3	US-08-978-741-4	Sequence 4, Appli
4	14.4	90.0	1300	4	US-09-333-729A-6	Sequence 6, Appli
5	14.4	90.0	1514	3	US-08-978-741-1	Sequence 1, Appli
6	14.4	90.0	1514	4	US-09-333-729A-2	Sequence 2, Appli
7	14.4	90.0	5009	3	US-08-978-741-7	Sequence 7, Appli
8	14.4	90.0	5009	4	US-09-333-729A-8	Sequence 8, Appli
9	14.4	90.0	11284	3	US-08-978-741-5	Sequence 5, Appli
c 10	14	87.5	79	3	US-08-463-903-78	Sequence 78, Appl
c 11	14	87.5	79	4	US-07-935-695-78	Sequence 78, Appl
12	14	87.5	95	3	US-08-463-903-79	Sequence 79, Appl
13	14	87.5	95	4	US-07-935-695-79	Sequence 79, Appl
14	13.4	83.8	206	2	US-08-485-657A-4	Sequence 4, Appli
15	13.4	83.8	208	5	PCT-US95-02303-4	Sequence 4, Appli
c 16	13.4	83.8	559	4	US-08-975-762-4	Sequence 4, Appli
c 17	13.4	83.8	559	4	US-08-821-324-4	Sequence 4, Appli
c 18	13.4	83.8	559	4	US-09-295-028-4	Sequence 4, Appli
c 19	13.4	83.8	559	4	US-09-106-582-4	Sequence 4, Appli
c 20	13.4	83.8	1330	2	US-08-933-750C-80	Sequence 80, Appl
c 21	13.4	83.8	1330	3	US-09-234-613-80	Sequence 80, Appl
c 22	13.4	83.8	1656	1	US-08-324-465-2	Sequence 2, Appli
c 23	13.4	83.8	1656	2	US-08-465-981-2	Sequence 2, Appli
c 24	13.4	83.8	1656	5	PCT-US93-11915-2	Sequence 2, Appli
c 25	13.4	83.8	1725	1	US-08-324-465-5	Sequence 5, Appli
c 26	13.4	83.8	1725	2	US-08-465-981-5	Sequence 5, Appli
c 27	13.4	83.8	1725	5	PCT-US93-11915-5	Sequence 5, Appli
28	13.4	83.8	1821	1	US-07-803-622E-3	Sequence 3, Appli
c 29	13.4	83.8	1919	4	US-08-975-762-40	Sequence 40, Appl
c 30	13.4	83.8	1919	4	US-09-295-028-40	Sequence 40, Appl
c 31	13.4	83.8	1919	4	US-09-106-582-40	Sequence 40, Appl
c 32	13.4	83.8	2265	2	US-08-940-332-1	Sequence 1, Appli
33	13.4	83.8	6263	2	US-08-781-802-3	Sequence 3, Appli
34	13.4	83.8	6263	4	US-08-694-078-3	Sequence 3, Appli
35	13.4	83.8	8779	2	US-08-750-703-4	Sequence 4, Appli
36	13.4	83.8	12752	2	US-08-459-146-1	Sequence 1, Appli
37	13.4	83.8	12752	2	US-08-459-065-1	Sequence 1, Appli
38	13	81.2	1814	2	US-08-483-151-1	Sequence 1, Appli
39	13	81.2	1814	5	PCT-US96-06427-1	Sequence 1, Appli
40	13	81.2	3116	4	US-09-362-831-10	Sequence 10, Appl
41	13	81.2	3624	1	US-07-951-715A-6	Sequence 6, Appli
42	13	81.2	3624	2	US-08-459-448A-6	Sequence 6, Appli
43	13	81.2	3624	3	US-08-459-595A-6	Sequence 6, Appli
44	13	81.2	3624	3	US-08-459-504B-6	Sequence 6, Appli
45	13	81.2	3624	3	US-08-459-444-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1100 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-16


  Query Match             90.0%;  Score 14.4;  DB 3;  Length 1100;
  Best Local Similarity   93.8%;  Pred. No. 51;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
             ||||||||||||| ||
Db     166 CCTTTCACCAACCTCC 181
```

```
RESULT   2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;   LENGTH: 1100
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-16


  Query Match             90.0%;  Score 14.4;  DB 4;  Length 1100;
  Best Local Similarity   93.8%;  Pred. No. 51;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    166 cctttcaccaacctcc 181


RESULT   3
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-4


  Query Match             90.0%;  Score 14.4;  DB 3;  Length 1300;
  Best Local Similarity   93.8%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    301 CCTTTCACCAACCTCC 316


RESULT   4
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 6
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6


  Query Match             90.0%;  Score 14.4;  DB 4;  Length 1300;
  Best Local Similarity   93.8%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    301 cctttcaccaacctcc 316


RESULT   5
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
```

```
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1


  Query Match             90.0%;  Score 14.4;  DB 3;  Length 1514;
  Best Local Similarity   93.8%;  Pred. No. 53;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    166 CCTTTCACCAACCTCC 181


RESULT   6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 2
;   LENGTH: 1514
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-2


  Query Match             90.0%;  Score 14.4;  DB 4;  Length 1514;
  Best Local Similarity   93.8%;  Pred. No. 53;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    166 cctttcaccaacctcc 181


RESULT   7
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5009 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-7


  Query Match             90.0%;  Score 14.4;  DB 3;  Length 5009;
  Best Local Similarity   93.8%;  Pred. No. 61;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    101 CCTTTCACCAACCTCC 116


RESULT   8
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 8
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-8


  Query Match             90.0%;  Score 14.4;  DB 4;  Length 5009;
  Best Local Similarity   93.8%;  Pred. No. 61;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cctttcaccaaccccc 16
          ||||||||||||| ||
Db    101 cctttcaccaacctcc 116


RESULT   9
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
```

```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/978,741
;       FILING DATE:  26-No. 6100076-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/792498
;       FILING DATE:  31
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Svoboda, Craig G.
;       REGISTRATION NUMBER:  39,044
;       REFERENCE/DOCKET NUMBER:  P1041P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650/225-1489
;       TELEFAX:  650/952-9881
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11284 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-978-741-5


  Query Match             90.0%;  Score 14.4;  DB 3;  Length 11284;
  Best Local Similarity   93.8%;  Pred. No. 66;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaaccccc 16
           ||||||||||||| ||
Db    4401 CCTTTCACCAACCTCC 4416


RESULT  10
US-08-463-903-78/c
; Sequence 78, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
;  APPLICANT: Mezes, Peter S.
;  APPLICANT: Richard, Ruth A.
;  APPLICANT: Affholter, Joseph A.
;  APPLICANT: Kotite, Nicolas J.
;  TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE: 40224A US
;  CURRENT APPLICATION NUMBER: US/08/463,903
;  CURRENT FILING DATE: 1995-06-05
;  EARLIER APPLICATION NUMBER: US 07/935,695
;  EARLIER FILING DATE: 1992-08-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 78
;   LENGTH: 79
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: DRb-3AH5' primer
;   LOCATION: 1..79
US-08-463-903-78


  Query Match             87.5%;  Score 14;  DB 3;  Length 79;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaacccc 15
           ||||||||||||||
Db      68 CTTTCACCAACCCC 55


RESULT  11
US-07-935-695-78/c
; Sequence 78, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
;  APPLICANT:  Mezes, Peter S.
;  APPLICANT:  Richard, Ruth A.
;  APPLICANT:  Affholter, Joseph A.
;  APPLICANT:  Kotite, Nicolas J.
;  TITLE OF INVENTION:  Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE:  40224A US
;  CURRENT APPLICATION NUMBER: US/07/935,695
;  CURRENT FILING DATE:  1992-08-21
;  PRIOR APPLICATION NUMBER:  US 08/463,903
;  PRIOR FILING DATE:  1995-06-05
;  NUMBER OF SEQ ID NOS:  102
;  SOFTWARE:  MS-Word for Windows, Ver. 7.0
; SEQ ID NO 78
;   LENGTH: 79
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: DRb-3AH5' primer
;   LOCATION: 1..79
;   OTHER INFORMATION: :
US-07-935-695-78


  Query Match             87.5%;  Score 14;  DB 4;  Length 79;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaacccc 15
           ||||||||||||||
Db      68 CTTTCACCAACCCC 55


RESULT  12
US-08-463-903-79
; Sequence 79, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
;  APPLICANT: Mezes, Peter S.
;  APPLICANT: Richard, Ruth A.
;  APPLICANT: Affholter, Joseph A.
;  APPLICANT: Kotite, Nicolas J.
;  TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE: 40224A US
;  CURRENT APPLICATION NUMBER: US/08/463,903
;  CURRENT FILING DATE: 1995-06-05
;  EARLIER APPLICATION NUMBER: US 07/935,695
;  EARLIER FILING DATE: 1992-08-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 79
;   LENGTH: 95
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: DRb-3AHV3' primer
;   LOCATION: 1..95
US-08-463-903-79


  Query Match             87.5%;  Score 14;  DB 3;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaacccc 15
           ||||||||||||||
```

```
Db      31 ctttcaccaacccc 44


RESULT  13
US-07-935-695-79
; Sequence 79, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
;  APPLICANT:  Mezes, Peter S.
;  APPLICANT:  Richard, Ruth A.
;  APPLICANT:  Affholter, Joseph A.
;  APPLICANT:  Kotite, Nicolas J.
;  TITLE OF INVENTION:  Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE:  40224A US
;  CURRENT APPLICATION NUMBER: US/07/935,695
;  CURRENT FILING DATE:  1992-08-21
;  PRIOR APPLICATION NUMBER:  US 08/463,903
;  PRIOR FILING DATE:  1995-06-05
;  NUMBER OF SEQ ID NOS:  102
;  SOFTWARE:  MS-Word for Windows, Ver. 7.0
; SEQ ID NO 79
;   LENGTH: 95
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: DRb-3AHV3' primer
;   LOCATION: 1..95
;   OTHER INFORMATION: :
US-07-935-695-79


  Query Match             87.5%;  Score 14;  DB 4;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ctttcaccaacccc 15
           ||||||||||||||
Db      31 ctttcaccaacccc 44


RESULT  14
US-08-485-657A-4
; Sequence 4, Application US/08485657A
; Patent No. 5942389
;  GENERAL INFORMATION:
;    APPLICANT:  Kirschling, Deborah J
;    APPLICANT:  Gudkov, Andrei
;    APPLICANT:  Roninson, Igor B
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive, 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,657A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5942389nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  93,354-N
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  206 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-485-657A-4


  Query Match             83.8%;  Score 13.4;  DB 2;  Length 206;
  Best Local Similarity   93.3%;  Pred. No. 1.4e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           | ||||||||||||||
Db     129 CATTTCACCAACCCC 143


RESULT  15
PCT-US95-02303-4
; Sequence 4, Application PC/TUS9502303
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02303
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US95-02303-4


  Query Match             83.8%;  Score 13.4;  DB 5;  Length 208;
  Best Local Similarity   93.3%;  Pred. No. 1.4e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cctttcaccaacccc 15
           | ||||||||||||||
Db     130 CATTTCACCAACCCC 144


Search completed: July 29, 2002, 23:56:13
Job time: 4754 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:55:06 ; Search time 1921.77 Seconds
(without alignments)
119.781 Million cell updates/sec

Title: US-09-530-663B-17
Perfect score: 11
Sequence: 1 tctcaccaacc 11

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	100.0	12	6	I24588	I24588 Sequence 16
c 2	11	100.0	16	6	I27905	I27905 Sequence 77
c 3	11	100.0	16	6	I27996	I27996 Sequence 16
c 4	11	100.0	17	6	I28051	I28051 Sequence 22
c 5	11	100.0	19	6	I28019	I28019 Sequence 19
6	11	100.0	24	6	A42460	A42460 Sequence 9
c 7	11	100.0	25	6	AR157073	AR157073 Sequence
c 8	11	100.0	27	6	AX128308	AX128308 Sequence
c 9	11	100.0	39	6	AX044055	AX044055 Sequence
c 10	11	100.0	39	6	AX044109	AX044109 Sequence
c 11	11	100.0	39	6	AX044157	AX044157 Sequence
c 12	11	100.0	54	6	AR040506	AR040506 Sequence
c 13	11	100.0	69	6	AR097013	AR097013 Sequence
14	11	100.0	74	6	AR097014	AR097014 Sequence
c 15	11	100.0	131	6	E15304	E15304 Oryza sativ
c 16	11	100.0	152	9	HUMMHDHOB	M30182 Human MHC c
17	11	100.0	154	10	S57440S05	S57459 lipocortin
c 18	11	100.0	156	11	AU025871	AU025871 Rattus no
c 19	11	100.0	175	9	S79786	S79786 HLA-DRB1 (D
20	11	100.0	209	6	AX237100	AX237100 Sequence
c 21	11	100.0	213	9	AF406781	AF406781 Homo sapi
22	11	100.0	220	6	AR034255	AR034255 Sequence
c 23	11	100.0	225	9	HUMMHDR413	M37771 Human MHC c
c 24	11	100.0	225	9	HUMMHDR414	M37770 Human MHC c
c 25	11	100.0	225	9	HUMMHDR4B1	M64794 Human MHC c
c 26	11	100.0	227	9	HSDRB1125	Z50746 H.sapiens D
c 27	11	100.0	228	9	HSU25442	U25442 Human MHC c
c 28	11	100.0	228	9	HSU25639	U25639 Human MHC c
c 29	11	100.0	231	9	HSDNAHLAB	X75347 H.sapiens D
c 30	11	100.0	231	9	HSDRB1	X70788 H.sapiens H
31	11	100.0	233	8	AY022841	AY022841 Oryza sat
c 32	11	100.0	234	9	HSAJ3124	AJ223124 Homo sapi
c 33	11	100.0	234	9	AF028588	AF028588 Homo sapi
c 34	11	100.0	234	9	AF028589	AF028589 Homo sapi
c 35	11	100.0	234	9	AF028590	AF028590 Homo sapi
c 36	11	100.0	234	9	AF144105	AF144105 Homo sapi
c 37	11	100.0	234	9	AF152844	AF152844 Homo sapi
c 38	11	100.0	234	9	AF152845	AF152845 Homo sapi
c 39	11	100.0	234	9	AF278701	AF278701 Homo sapi
c 40	11	100.0	235	9	HSDRB3MOB	X99771 H.sapiens H
c 41	11	100.0	236	9	AF093411	AF093411 Homo sapi
c 42	11	100.0	237	9	HSDRB13EX	X93315 H.sapiens H
c 43	11	100.0	237	9	HUMDRB4X	L21985 Human major
c 44	11	100.0	237	9	HUMSEQA	L26306 Human DNA s
c 45	11	100.0	237	11	G19684	G19684 human STS A

ALIGNMENTS

10 of 16

RESULT 1
I24588
LOCUS I24588 12 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 16 from patent US 5545526.
ACCESSION I24588
VERSION I24588.1 GI:1604458
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 12)
AUTHORS Baxter-Lowe,L.Ann.
TITLE Method for HLA Typing
JOURNAL Patent: US 5545526-A 16 13-AUG-1996;
FEATURES Location/Qualifiers
source 1. .12
/organism="unknown"
BASE COUNT 3 a 7 c 0 g 2 t
ORIGIN

Query Match 100.0%; Score 11; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.7e+04;

```
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db      2 TCTCACCAACC 12


RESULT   2
I27905/c
LOCUS       I27905                   16 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 77 from patent US 5567809.
ACCESSION   I27905
VERSION     I27905.1  GI:1818681
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.
  TITLE     Methods and reagents for HLA DRbeta DNA typing
  JOURNAL   Patent: US 5567809-A 77 22-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        2 a      2 c      8 g      4 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 5.4e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     13 TCTCACCAACC 3


RESULT   3
I27996/c
LOCUS       I27996                   16 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 168 from patent US 5567809.
ACCESSION   I27996
VERSION     I27996.1  GI:1818772
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.
  TITLE     Methods and reagents for HLA DRbeta DNA typing
  JOURNAL   Patent: US 5567809-A 168 22-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        2 a      1 c      9 g      4 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 5.4e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     13 TCTCACCAACC 3


RESULT   4
I28051/c
LOCUS       I28051                   17 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 223 from patent US 5567809.
ACCESSION   I28051
VERSION     I28051.1  GI:1818827
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.
  TITLE     Methods and reagents for HLA DRbeta DNA typing
  JOURNAL   Patent: US 5567809-A 223 22-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        2 a      2 c      9 g      4 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 5.4e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     14 TCTCACCAACC 4


RESULT   5
I28019/c
LOCUS       I28019                   19 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 191 from patent US 5567809.
ACCESSION   I28019
VERSION     I28019.1  GI:1818795
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Apple,R.J., Erlich,H.A., Griffith,R.L. and Scharf,S.J.
  TITLE     Methods and reagents for HLA DRbeta DNA typing
  JOURNAL   Patent: US 5567809-A 191 22-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="unknown"
BASE COUNT        3 a      2 c      9 g      5 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 5.3e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     15 TCTCACCAACC 5


RESULT   6
A42460
LOCUS       A42460                   24 bp    DNA     linear   PAT 06-MAR-1997
DEFINITION  Sequence 9 from Patent WO9503331.
ACCESSION   A42460
VERSION     A42460.1  GI:2297917
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Whittaker,J.L. and Morten,J.E.
  TITLE     HUMAN MHC CLASS II DOUBLE TRANSGENE AND USES
  JOURNAL   Patent: WO 9503331-A 9 02-FEB-1995;
            ZENECA LTD (GB)
COMMENT     Other publication AU 7231494 950220
            Other publication GB 2280186 950125.
FEATURES             Location/Qualifiers
     source          1. .24
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        6 a      10 c       4 g       4 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 5e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     11 TCTCACCAACC 21


RESULT   7
AR157073/c
LOCUS       AR157073                  25 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 10 from patent US 6242588.
ACCESSION   AR157073
VERSION     AR157073.1  GI:15125777
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Sheppard,P.O., Piddington,C.S. and Ellsworth,J.L.
  TITLE     Testis specific glycoprotein zpep10
  JOURNAL   Patent: US 6242588-A 10 05-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="unknown"
BASE COUNT        4 a      5 c       7 g       9 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     22 TCTCACCAACC 12


RESULT   8
AX128308/c
LOCUS       AX128308                  27 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 75 from Patent WO0130992.
ACCESSION   AX128308
VERSION     AX128308.1  GI:14134829
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Koike,C.
  TITLE     _g(a)1-3 galactosyltransferase gene and promoter
  JOURNAL   Patent: WO 0130992-A 75 03-MAY-2001;
            UNIV. PITTSBURGH OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION
            (US)
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer for identifying murine exons 6 and 7"
BASE COUNT        7 a      4 c      10 g       6 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     18 TCTCACCAACC 8


RESULT   9
AX044055/c
LOCUS       AX044055                  39 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 9 from Patent WO0066748.
ACCESSION   AX044055
VERSION     AX044055.1  GI:11342933
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Hawkes,T.R., Warner,S.A., Andrews,C.J., Bachoo,S. and
            Pickerill,A.P.
  TITLE     Herbicide resistant plants
  JOURNAL   Patent: WO 0066748-A 9 09-NOV-2000;
            ZENECA LIMITED (GB)
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT        7 a      9 c      15 g       8 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     29 TCTCACCAACC 19


RESULT  10
AX044109/c
LOCUS       AX044109                  39 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 9 from Patent WO0066747.
ACCESSION   AX044109
VERSION     AX044109.1  GI:11342987
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Hawkes,T.R., Warner,S.A., Andrews,C.J., Bachoo,S. and
            Pickerill,A.P.
  TITLE     Herbicide resistant plants
  JOURNAL   Patent: WO 0066747-A 9 09-NOV-2000;
            ZENECA LIMITED (GB)
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT        7 a      9 c      15 g       8 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     29 TCTCACCAACC 19
```

```
RESULT  11
AX044157/c
LOCUS       AX044157                  39 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 9 from Patent WO0066746.
ACCESSION   AX044157
VERSION     AX044157.1  GI:11343035
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 39)
  AUTHORS   Hawkes,T.R., Warner,S.A., Andrews,C.J., Bachoo,S. and
            Pickerill,A.P.
  TITLE     Herbicide resistant plants
  JOURNAL   Patent: WO 0066746-A 9 09-NOV-2000;
            ZENECA LIMITED (GB)
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Primer"
BASE COUNT        7 a      9 c     15 g      8 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 39;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     29 TCTCACCAACC 19


RESULT  12
AR040506/c
LOCUS       AR040506                  54 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1354 from patent US 5807743.
ACCESSION   AR040506
VERSION     AR040506.1  GI:5959869
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 54)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE     Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 1354 15-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .54
                     /organism="unknown"
BASE COUNT       15 a      9 c     18 g     12 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 54;
  Best Local Similarity  100.0%;  Pred. No. 4.3e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     11 TCTCACCAACC 1


RESULT  13
AR097013/c
LOCUS       AR097013                  69 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 76 from patent US 6071515.
ACCESSION   AR097013
VERSION     AR097013.1  GI:12805743
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Mezes,P.S., Richard,R.A., Affholter,J.A. and Kotite,N.J.
  TITLE     Dimer and multimer forms of single chain polypeptides
  JOURNAL   Patent: US 6071515-A 76 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="unknown"
BASE COUNT       16 a     16 c     24 g     13 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 69;
  Best Local Similarity  100.0%;  Pred. No. 4.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     41 TCTCACCAACC 31


RESULT  14
AR097014
LOCUS       AR097014                  74 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 77 from patent US 6071515.
ACCESSION   AR097014
VERSION     AR097014.1  GI:12805744
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 74)
  AUTHORS   Mezes,P.S., Richard,R.A., Affholter,J.A. and Kotite,N.J.
  TITLE     Dimer and multimer forms of single chain polypeptides
  JOURNAL   Patent: US 6071515-A 77 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .74
                     /organism="unknown"
BASE COUNT       14 a     25 c     17 g     18 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 6;  Length 74;
  Best Local Similarity  100.0%;  Pred. No. 4.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     32 TCTCACCAACC 42


RESULT  15
E15304/c
LOCUS       E15304                   131 bp    DNA     linear   PAT 28-JUL-1999
DEFINITION  Oryza sativa microsatellite marker.
ACCESSION   E15304
VERSION     E15304.1  GI:5709987
KEYWORDS    JP 1998057073-A/28.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 131)
  AUTHORS   Akagi,H., Fujimura,T., Yokozeki,S. and Inagaki,A.
  TITLE     MICROSATTELITE MARKER FOR DISCRIMINATING RICE PLANT CULTIVAR AND
            TEST FOR PURITY OF PLANT SEED
  JOURNAL   Patent: JP 1998057073-A 28 03-MAR-1998;
            MITSUI PETROCHEM IND LTD
COMMENT     OS   Oryza sativa
            PN   JP 1998057073-A/28
            PD   03-MAR-1998
```

```
            PF   25-FEB-1997 JP 1997040226
            PR   13-JUN-1996 JP 96P   152657
            PI   AKAGI HIROMORI, FUJIMURA TATSUTO, YOKOZEKI SUKEYOSHI, PI
            INAGAKI AKIKO
            PC   C12N15/09,C12Q1/68;
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .131
            FT                   /organism='Oryza sativa'
            FT   repeat_region   1. .131
            FT                   /rpt_family='microsatellite'.
FEATURES             Location/Qualifiers
     source          1. .131
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
BASE COUNT       26 a     27 c     63 g     15 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 6;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||||
Db      41 TCTCACCAACC 31


Search completed: July 29, 2002, 23:55:08
Job time: 7024 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 00:01:19 ; Search time 285.14 Seconds
(without alignments)
66.234 Million cell updates/sec

Title: US-09-530-663B-17
Perfect score: 11
Sequence: 1 tctcaccaacc 11

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	11	100.0	12	17	AAX79382	HLA-DR typing prob
	2	11	100.0	12	17	AAT41819	HLA allele, HLA-DR
c	3	11	100.0	12	19	AAV16578	Probe G86 used to
c	4	11	100.0	12	22	AAF92638	HLA-DR typing prob
c	5	11	100.0	16	13	AAQ26124	HLA-DR beta sub-ty
c	6	11	100.0	16	13	AAQ26196	HLA-DR beta sub-ty
c	7	11	100.0	16	24	ABL31303	Human HLA genotypi
c	8	11	100.0	17	13	AAQ26251	HLA-DR beta sub-ty
c	9	11	100.0	17	24	ABL30817	Human HLA genotypi
c	10	11	100.0	18	22	AAS01746	Glucanase genomic
c	11	11	100.0	18	24	ABL30757	Human HLA genotypi
c	12	11	100.0	18	24	ABL30790	Human HLA genotypi
c	13	11	100.0	19	11	AAQ06439	Oligonucleotide pr
c	14	11	100.0	19	13	AAQ26219	HLA-DR beta sub-ty
c	15	11	100.0	19	17	AAX79408	HLA-DR typing prob
c	16	11	100.0	19	19	AAV16604	Probe VAL86 used t
c	17	11	100.0	19	22	AAF92664	HLA-DR typing prob
	18	11	100.0	22	21	AAC66594	Gene typing PCR pr
	19	11	100.0	24	16	AAQ99480	Oligonucleotide pr
c	20	11	100.0	25	21	AAA49659	Human testis speci
c	21	11	100.0	25	22	AAD07332	Human testis-speci
c	22	11	100.0	27	22	AAS05364	Mouse alpha-1,3 ga
c	23	11	100.0	39	21	AAC87155	Rice EPSPS PCR pri
c	24	11	100.0	39	21	AAC88374	Primer G1. Synthe
c	25	11	100.0	39	21	AAC89290	Primer G1. Synthe
c	26	11	100.0	54	19	AAV94562	Canine IL-2 recept
	27	11	100.0	122	21	AAC18184	Human secreted pro
c	28	11	100.0	131	19	AAV16156	Microsatellite mar
c	29	11	100.0	135	22	AAF92615	HLA-DR beta allele
c	30	11	100.0	170	21	AAC04125	Human secreted pro
c	31	11	100.0	203	22	ABA70729	Human foetal liver
c	32	11	100.0	203	22	AAK18980	Human brain expres
c	33	11	100.0	203	22	AAK44930	Human bone marrow
c	34	11	100.0	203	22	AAI50900	Probe #19586 used
	35	11	100.0	209	22	AAK54351	Human haematologic
	36	11	100.0	220	16	AAT04587	Spleen necrosis vi
	37	11	100.0	220	21	AAZ51114	Eucaryotic gene ex
c	38	11	100.0	240	17	AAX79459	HLA-DR beta allele
c	39	11	100.0	240	22	AAF92582	HLA-DR beta allele
	40	11	100.0	264	21	AAC50284	Arabidopsis thalia
c	41	11	100.0	268	13	AAQ25135	DR1. Synthetic.
c	42	11	100.0	268	20	AAZ39269	HLA allele DRB1*04
c	43	11	100.0	269	13	AAQ26358	DRB1*1105 allele.
c	44	11	100.0	269	13	AAQ26366	DRB1*1305 allele.
c	45	11	100.0	269	17	AAX79486	HLA-DR beta allele

ALIGNMENTS

RESULT 1
AAX79382/c
ID AAX79382 standard; DNA; 12 BP.
XX
AC AAX79382;
XX
DT 17-AUG-1999 (first entry)
XX
DE HLA-DR typing probe G86.
XX
KW Tissue typing; human leukocyte antigen; HLA; MHC; donor; allele; PCR;
KW major histocompatibility complex; bone marrow transplant; primer;
KW amplification; polymerase chain reaction; probe; polymorphism;
KW sequence-specific oligonucleotide probe hybridisation; ss.
XX
OS Synthetic.
XX
PN US5468611-A.
XX
PD 21-NOV-1995.
XX
PF 08-APR-1993; 93US-0045530.
XX
PR 27-JUN-1990; 90US-0544218.
PR 08-APR-1993; 93US-0045530.
XX
PA (BLOO-) BLOOD CENT RES FOUND INC.
XX
PI Baxter-Lowe LA, Gorski JA;
XX
DR WPI; 1996-010091/01.
XX

PT Improved method for HLA typing - by DNA amplification and
PT sequence-specific oligo:nucleotide hybridisation, used to select
PT bone marrow donors
XX
PS Disclosure; Column 19-20; 20pp; English.
XX
CC A novel method of typing the human leukocyte antigen (HLA) of the major
CC histocompatibility complex (MHC), esp. for typing donors for bone marrow
CC transplants, involves determining if the donor tissue HLA-DR alleles are
CC selected from the gp.: HLA-DRW52C, DR12a,b, DR3a,n, DR5a-e, DRNew1,
CC DR6a, DR8a-d, DRW53a-c, DR4a-f, DR7, DR9, DR2a-c B3, DR2a-d B1, DR10 and
CC DR1a-c. The method uses PCR to amplify these regions followed by
CC sequence-specific oligonucleotide probe hybridisation (SSOPH) using the
CC probes AAX79365-X79429. SSOPH allows detection of polymorphisms that
CC predict differences at a single amino acid level thus reducing errors
CC and improving the chance of successfully matching tissues.
XX
SQ Sequence 12 BP; 2 A; 0 C; 7 G; 3 T; 0 other;

Query Match 100.0%; Score 11; DB 17; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||
Db 11 TCTCACCAACC 1

RESULT 2
AAT41819
ID AAT41819 standard; DNA; 12 BP.
XX
AC AAT41819;
XX
DT 18-DEC-1996 (first entry)
XX
DE HLA allele, HLA-DRB1*08, *12 and *1404 resolution probe, G86.
XX
KW Human leukocyte antigen; HLA; allele; HLA-DR*08; HLA-DR*12; locus B1;
KW polymorphism; amplify; conserved region; detection; primer; probe;
KW tissue matching; identifying disease susceptibility; ss.
XX
OS Synthetic.
XX
PN US5545526-A.
XX
PD 13-AUG-1996.
XX
PF 27-JUN-1990; 90US-0544218.
XX
PR 01-MAR-1993; 93US-0025038.
PR 27-JUN-1990; 90US-0544218.
XX
PA (BLOO-) BLOOD CENT RES FOUND INC.
XX
PI Baxter-Lowe LA;
XX
DR WPI; 1996-383664/38.
XX
PT Human leukocyte antigen typing of tissue samples - using
PT allele-specific amplification to distinguish allele pairs
XX
PS Example 1; Column 19; 24pp; English.
XX
CC The sequences given in AAT41811-20 represent probes which were used to
CC resolve the human leukocyte antigen (HLA) DRB1 alleles, DRB1*08, *12
CC and *1404. This probe sequence hybridises to the Gly86 coding region
CC found in alleles *0801, *0802, *0803 and 0805. These probes may be used
CC in the method of invention which concerns HLA typing of a sample for an
CC unknown pair of alleles. The pair of alleles comprises one of two known
CC types which have the same overall set of polymorphisms but have a
CC different distribution of polymorphisms between their two alleles. The
CC method comprises selectively amplifying the DNA of just one allele of
CC the unknown pair and analysing the amplified DNA to determine which
CC polymorphisms are present in that allele, and therefore assigning the
CC unknown pair to the known type having that allele. The method comprises
CC three test stages. The first stage is to establish the number of
CC alleles present in each sample. Primers corresponding to fairly well
CC conserved regions of a locus will increase the likelihood that unknown
CC alleles will be amplified and potentially detected by hybridisation with
CC a probe. In the second stage, the group or basic type identified
CC determines which set of allele specific primers will be used. The first
CC of the two primers comprises an opt. labeled sequence common to each
CC allele of the group identified in the first stage but different from
CC other groups identified in stage one. The second primer may be a
CC mixture of different labeled primers, complementary to two or more
CC sequences within the group, or the amplification may be performed with
CC only one second primer to detect the presence of a single group of
CC alleles. In the third stage the specific allele is determined. This
CC may be done by amplification or hybridisation using a radiolabelled
CC probe. The method may be used for tissue matching, identifying disease
CC susceptibility, etc. The method of the invention esp. distinguishes
CC between DQB1*0304/DQB1*03032 and DQB1*0301/DQB1*0302.
XX
SQ Sequence 12 BP; 3 A; 7 C; 0 G; 2 T; 0 other;

Query Match 100.0%; Score 11; DB 17; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||
Db 2 tctcaccaacc 12

RESULT 3
AAV16578/c
ID AAV16578 standard; DNA; 12 BP.
XX
AC AAV16578;
XX
DT 12-JUN-1998 (first entry)
XX
DE Probe G86 used to identify HLA-DR sequences.
XX
KW DR region; major histocompatibility complex; HLA-DR; HLA-typing;
KW HLA-DR beta consensus sequence; allelic polymorphism;
KW HLA-DR beta-allelic polymorphism; probe; bone marrow; transplant; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US5702885-A.
XX
PD 30-DEC-1997.
XX
PF 08-APR-1993; 93US-0057957.
XX
PR 27-JUN-1990; 90US-0544218.
XX
PA (BLOO-) BLOOD CENT RES FOUND INC.
XX
PI Baxter-Lowe LA, Gorski JA;
XX
DR WPI; 1998-076408/07.
XX
PT Oligo:nucleotide probes and primers and methods for HLA typing -
PT particularly for tissue typing for bone marrow transplants
XX
PS Disclosure; Column 19; 20pp; English.
XX
CC Probes AAV16561-624 are used to identify differences in the DR region of
CC human major histocompatibility complex (HLA-DR). The specification
CC describes a method for HLA-typing, which includes an oligonucleotide

```
CC   probe which undergoes sequence-specific hybridisation with an HLA-DR
CC   beta consensus sequence at positions 61-64. The probe contains a
CC   labelling substance other than a nucleotide sequence, which facilitates
CC   detection of the probe. The HLA sequence of a subject is PCR amplified,
CC   and a probe that recognises an allelic polymorphism at a selected HLA
CC   locus is contacted with the amplified product. This first probe
CC   recognises a HLA-DR beta-allelic polymorphism. A second (different)
CC   probe is brought into contact with a second sample of the amplified DNA
CC   in a separate reaction, and hybridisation detected. The probes and
CC   primers are used for HLA typing, e.g. for tissue, especially bone
CC   marrow, transplants.
XX
SQ   Sequence 12 BP; 2 A; 0 C; 7 G; 3 T; 0 other;


  Query Match            100.0%;  Score 11;  DB 19;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     11 TCTCACCAACC 1


RESULT   4
AAF92638/c
ID   AAF92638 standard; DNA; 12 BP.
XX
AC   AAF92638;
XX
DT   16-MAY-2001  (first entry)
XX
DE   HLA-DR typing probe #18.
XX
KW   Human; leukocyte antigen; HLA; typing; sequence specific probe;
KW   SSOPH; ss.
XX
OS   Homo sapiens.
XX
PN   US6194147-B1.
XX
PD   27-FEB-2001.
XX
PF   30-DEC-1997;   97US-0000805.
XX
PR   27-JUN-1990;   90US-0544218.
PR   08-APR-1993;   93US-0057957.
XX
PA   (BLOO-) BLOOD CENT RES FOUND INC.
XX
PI   Baxter-Lowe LA,  Gorski JA;
XX
DR   WPI; 2001-217923/22.
XX
PT   Human leukocyte antigen typing by amplifying a sample followed by
PT   sequence specific oligonucleotide hybridization with labeled
PT   oligonucleotide probes that hybridize with a series of known control
PT   DNA sequences -
XX
PS   Disclosure; Column 11-14; 16pp; English.
XX
CC   The present invention relates to human leukocyte antigen (HLA) typing.
CC   The method involves detecting polymorphic residues by sequence
CC   specific oligonucleotide probe hybridization (SSOPH) with labeled
CC   oligonucleotide probes.
XX
SQ   Sequence 12 BP; 2 A; 0 C; 7 G; 3 T; 0 other;


  Query Match            100.0%;  Score 11;  DB 22;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     11 TCTCACCAACC 1


RESULT   5
AAQ26124/c
ID   AAQ26124 standard; DNA; 16 BP.
XX
AC   AAQ26124;
XX
DT   04-JAN-1993  (first entry)
XX
DE   HLA-DR beta sub-type tailed probe DRB15 hybridising region.
XX
KW   Tissue typing; identity determination; disease susceptible; ss.
XX
OS   Synthetic.
XX
PN   WO9210589-A.
XX
PD   25-JUN-1992.
XX
PF   06-DEC-1991;   91WO-US09294.
XX
PR   06-DEC-1990;   90US-0623098.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Apple RJ,  Begovich AB,  Bugawan T,  Erlich HA,  Griffith RL;
PI   Scharf SJ;
XX
DR   WPI; 1992-234644/28.
XX
PT   Method for determining HLA-DR beta sub-type in DNA sample -
PT   comprises amplification and hybridisation with probes and
PT   primers, useful in tissue typing
XX
PS   Example; Page 37; 90pp; English.
XX
CC   The sequence is that of the hybridising region of tailed probe DRB15 for
CC   use in a method for determining HLA-DR beta sub-type in a nucleic acid
CC   sample. The method allows specific nucleic acid sequences of the second
CC   exon of HLA-DR beta genes to be amplified then probed for identification
CC   of polymorphic sequences. The amplified DNA is useful for typing
CC   homozgous or heterozygous samples from a variety of sources and for
CC   detecting allelic variants not distinguishable by serological methods.
CC   The typing system can be used in a reverse dot blot format which is
CC   simple and rapid to perform, produces detectable signals in minutes and
CC   can be utilised in tissue typing, determination of individual identity
CC   and identifying disease susceptible individuals. Preliminary testing
CC   shows that the probe is more preferred than others. The probe is
CC   used with the HRP-labelled, untailed probe CRX56.
CC   See also AAQ26092-Q26367.
XX
SQ   Sequence 16 BP; 2 A; 2 C; 8 G; 4 T; 0 other;


  Query Match            100.0%;  Score 11;  DB 13;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     13 TCTCACCAACC 3


RESULT   6
AAQ26196/c
ID   AAQ26196 standard; DNA; 16 BP.
XX
AC   AAQ26196;
XX
```

```
DT   04-JAN-1993  (first entry)
XX
DE   HLA-DR beta sub-type tailed probe DRB92 hybridising region.
XX
KW   Tissue typing; identity determination; disease susceptible; ss.
XX
OS   Synthetic.
XX
PN   WO9210589-A.
XX
PD   25-JUN-1992.
XX
PF   06-DEC-1991;   91WO-US09294.
XX
PR   06-DEC-1990;   90US-0623098.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Apple RJ,  Begovich AB,  Bugawan T,  Erlich HA,  Griffith RL;
PI   Scharf SJ;
XX
DR   WPI; 1992-234644/28.
XX
PT   Method for determining HLA-DR beta sub-type in DNA sample -
PT   comprises amplification and hybridisation with probes and
PT   primers, useful in tissue typing
XX
PS   Example; Page 39; 90pp; English.
XX
CC   The sequence is that of the hybridising region of tailed probe DRB92 for
CC   use in a method for determining HLA-DR beta sub-type in a nucleic acid
CC   sample. The method allows specific nucleic acid sequences of the second
CC   exon of HLA-DR beta genes to be amplified then probed for identification
CC   of polymorphic sequences. The amplified DNA is useful for typing
CC   homozgous or heterozygous samples from a variety of sources and for
CC   detecting allelic variants not distinguishable by serological methods.
CC   The typing system can be used in a reverse dot blot format which is
CC   simple and rapid to perform, produces detectable signals in minutes and
CC   can be utilised in tissue typing, determination of individual identity
CC   and identifying disease susceptible individuals.
CC   See also AAQ26092-Q26367.
XX
SQ   Sequence 16 BP; 2 A; 1 C; 9 G; 4 T; 0 other;


  Query Match             100.0%;  Score 11;  DB 13;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      13 TCTCACCAACC 3


RESULT   7
ABL31303/c
ID   ABL31303 standard; DNA; 16 BP.
XX
AC   ABL31303;
XX
DT   21-MAR-2002  (first entry)
XX
DE   Human HLA genotyping oligonucleotide SEQ ID NO 792.
XX
KW   Human; human leukocyte antigen; HLA; genotype; polymorphism;
KW   immunogenetic; transplantation; genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192572-A1.
XX
PD   06-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-JP04662.
XX
PR   01-JUN-2000; 2000JP-0164798.
XX
PA   (NISN ) NISSHINBO IND INC.
PA   (SYST-) SYSTEM RES INC.
XX
PI   Inoko H,  Kagiya T,  Ichihara T,  Matsumura Y,  Moriya S,  Nishida M;
XX
DR   WPI; 2002-122074/16.
XX
PT   Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes
PT   of individuals e.g. by determining immunogenetic differences when
PT   transplanting between them -
XX
PS   Claim 10; Page 243; 345pp; Japanese.
XX
CC   The invention relates to a typing kit for judging human leukocyte antigen
CC   (HLA) genotype of a sample by hybridising a substrate on which 10-24 base
CC   oligonucleotides (ABL30512-ABL31809) originating in the sequences of
CC   genes e.g. belonging to HLA class I antigens on human genome and
CC   containing gene polymorphisms as alloantigens have been immobilised as
CC   primers for amplification of cleaved nucleic acids relating to gene
CC   polymorphisms. The method is useful for judging HLA genotypes of
CC   individuals by determining immunogenetic differences before transplanting
CC   between them, providing genetic information to decide compatibility of
CC   organ and tissue for transplantation e.g. of bone marrow, kidney, liver,
CC   pancreas, Langerhans islet in pancreas and cornea, susceptibility
CC   diagnosis of genetic diseases and identifying individuals.
XX
SQ   Sequence 16 BP; 2 A; 1 C; 9 G; 4 T; 0 other;


  Query Match             100.0%;  Score 11;  DB 24;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      13 TCTCACCAACC 3


RESULT   8
AAQ26251/c
ID   AAQ26251 standard; DNA; 17 BP.
XX
AC   AAQ26251;
XX
DT   04-JAN-1993  (first entry)
XX
DE   HLA-DR beta sub-type tailed probe DRB147 hybridising region.
XX
KW   Tissue typing; identity determination; disease susceptible; ss.
XX
OS   Synthetic.
XX
PN   WO9210589-A.
XX
PD   25-JUN-1992.
XX
PF   06-DEC-1991;   91WO-US09294.
XX
PR   06-DEC-1990;   90US-0623098.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Apple RJ,  Begovich AB,  Bugawan T,  Erlich HA,  Griffith RL;
PI   Scharf SJ;
XX
DR   WPI; 1992-234644/28.
XX
PT   Method for determining HLA-DR beta sub-type in DNA sample -
PT   comprises amplification and hybridisation with probes and
```

PT primers, useful in tissue typing
XX
PS Example; Page 41; 90pp; English.
XX
CC The sequence is that of the hybridising region of tailed probe DRB147
CC for use in a method for determining HLA-DR beta sub-type in a nucleic
CC acid sample. The method allows specific nucleic acid sequences of the
CC second exon of HLA-DR beta genes to be amplified then probed for
CC identification of polymorphic sequences. The amplified DNA is useful for
CC typing homozgous or heterozygous samples from a variety of sources and
CC for detecting allelic variants not distinguishable by serological
CC methods. The typing system can be used in a reverse dot blot format which
CC is simple and rapid to perform, produces detectable signals in minutes
CC and can be utilised in tissue typing, determination of individual
CC identity and identifying disease susceptible individuals. It has not yet
CC been tested. See also AAQ26092-Q26367.
XX
SQ Sequence 17 BP; 2 A; 2 C; 9 G; 4 T; 0 other;

Query Match 100.0%; Score 11; DB 13; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||||
Db 14 TCTCACCAACC 4

RESULT 9
ABL30817/c
ID ABL30817 standard; DNA; 17 BP.
XX
AC ABL30817;
XX
DT 21-MAR-2002 (first entry)
XX
DE Human HLA genotyping oligonucleotide SEQ ID NO 306.
XX
KW Human; human leukocyte antigen; HLA; genotype; polymorphism;
KW immunogenetic; transplantation; genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO200192572-A1.
XX
PD 06-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-JP04662.
XX
PR 01-JUN-2000; 2000JP-0164798.
XX
PA (NISN) NISSHINBO IND INC.
PA (SYST-) SYSTEM RES INC.
XX
PI Inoko H, Kagiya T, Ichihara T, Matsumura Y, Moriya S, Nishida M;
XX
DR WPI; 2002-122074/16.
XX
PT Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes
PT of individuals e.g. by determining immunogenetic differences when
PT transplanting between them -
XX
PS Claim 10; Page 151; 345pp; Japanese.
XX
CC The invention relates to a typing kit for judging human leukocyte antigen
CC (HLA) genotype of a sample by hybridising a substrate on which 10-24 base
CC oligonucleotides (ABL30512-ABL31809) originating in the sequences of
CC genes e.g. belonging to HLA class I antigens on human genome and
CC containing gene polymorphisms as alloantigens have been immobilised as
CC primers for amplification of cleaved nucleic acids relating to gene
CC polymorphisms. The method is useful for judging HLA genotypes of
CC individuals by determining immunogenetic differences before transplanting
CC between them, providing genetic information to decide compatibility of
CC organ and tissue for transplantation e.g. of bone marrow, kidney, liver,
CC pancreas, Langerhans islet in pancreas and cornea, susceptibility
CC diagnosis of genetic diseases and identifying individuals.
XX
SQ Sequence 17 BP; 3 A; 2 C; 9 G; 3 T; 0 other;

Query Match 100.0%; Score 11; DB 24; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||||
Db 15 TCTCACCAACC 5

RESULT 10
AAS01746/c
ID AAS01746 standard; DNA; 18 BP.
XX
AC AAS01746;
XX
DT 12-SEP-2001 (first entry)
XX
DE Glucanase genomic DNA sequencing primer Glucgene-1.
XX
KW Glucanase; endochitinase; exochitinase; cell-wall degradation; fungus;
KW transgenic plant; plant pathogen; bacteria; seafood waste; shell; ss;
KW chitin; chemical modification; glucan; sequencing primer; Glucgene-1.
XX
OS Fusarium venenatum.
XX
PN WO200116353-A1.
XX
PD 08-MAR-2001.
XX
PF 30-AUG-2000; 2000WO-US23802.
XX
PR 30-AUG-1999; 99US-0151582.
PR 11-AUG-2000; 2000US-0224946.
PR 28-AUG-2000; 2000US-0649747.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (USDA) US SEC OF AGRIC.
XX
PI Okubara PA, Blechl AE, Hohn TM, Berka RM;
XX
DR WPI; 2001-218524/22.
XX
PT Fusarium nucleic acids encoding polypeptides having glucanase,
PT endochitinase or exochitinase activity, useful for producing transgenic
PT plants which are resistant to plant pathogens, particularly Fusarium
PT species -
XX
PS Disclosure; Page 80; 216pp; English.
XX
CC The sequence represents a sequencing primer for DNA encoding the Fusarium
CC fungal enzyme, glucanase. Glucanase, endochitinase and exochitinase
CC are polypeptides with cell-wall degrading activity, derived from Fusarium
CC fungal genes. The associated nucleic acids can be used to produce
CC transgenic plants which are resistant to plant pathogens, particularly
CC Fusarium species. They can also be used to isolate homologous genes from
CC fungi to obtain genes which protect host cells, including fungi, bacteria
CC and plants against related fungal pathogens. The polypeptides, especially
CC chitinases and glucanases, are useful for degrading seafood waste, such
CC as shells that contain chitin, or for chemical modification of chitin or
CC glucan.
XX
SQ Sequence 18 BP; 3 A; 2 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 22; Length 18;

```
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||
Db      17 TCTCACCAACC 7


RESULT  11
ABL30757/c
ID   ABL30757 standard; DNA; 18 BP.
XX
AC   ABL30757;
XX
DT   21-MAR-2002  (first entry)
XX
DE   Human HLA genotyping oligonucleotide SEQ ID NO 246.
XX
KW   Human; human leukocyte antigen; HLA; genotype; polymorphism;
KW   immunogenetic; transplantation; genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192572-A1.
XX
PD   06-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-JP04662.
XX
PR   01-JUN-2000; 2000JP-0164798.
XX
PA   (NISN ) NISSHINBO IND INC.
PA   (SYST-) SYSTEM RES INC.
XX
PI   Inoko H,  Kagiya T,  Ichihara T,  Matsumura Y,  Moriya S,  Nishida M;
XX
DR   WPI; 2002-122074/16.
XX
PT   Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes
PT   of individuals e.g. by determining immunogenetic differences when
PT   transplanting between them -
XX
PS   Claim 10; Page 139; 345pp; Japanese.
XX
CC   The invention relates to a typing kit for judging human leukocyte antigen
CC   (HLA) genotype of a sample by hybridising a substrate on which 10-24 base
CC   oligonucleotides (ABL30512-ABL31809) originating in the sequences of
CC   genes e.g. belonging to HLA class I antigens on human genome and
CC   containing gene polymorphisms as alloantigens have been immobilised as
CC   primers for amplification of cleaved nucleic acids relating to gene
CC   polymorphisms. The method is useful for judging HLA genotypes of
CC   individuals by determining immunogenetic differences before transplanting
CC   between them, providing genetic information to decide compatibility of
CC   organ and tissue for transplantation e.g. of bone marrow, kidney, liver,
CC   pancreas, Langerhans islet in pancreas and cornea, susceptibility
CC   diagnosis of genetic diseases and identifying individuals.
XX
SQ   Sequence 18 BP; 3 A; 2 C; 9 G; 4 T; 0 other;


  Query Match             100.0%;  Score 11;  DB 24;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||
Db      15 TCTCACCAACC 5


RESULT  12
ABL30790/c
ID   ABL30790 standard; DNA; 18 BP.
XX
AC   ABL30790;
XX
DT   21-MAR-2002  (first entry)
XX
DE   Human HLA genotyping oligonucleotide SEQ ID NO 279.
XX
KW   Human; human leukocyte antigen; HLA; genotype; polymorphism;
KW   immunogenetic; transplantation; genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192572-A1.
XX
PD   06-DEC-2001.
XX
PF   01-JUN-2001; 2001WO-JP04662.
XX
PR   01-JUN-2000; 2000JP-0164798.
XX
PA   (NISN ) NISSHINBO IND INC.
PA   (SYST-) SYSTEM RES INC.
XX
PI   Inoko H,  Kagiya T,  Ichihara T,  Matsumura Y,  Moriya S,  Nishida M;
XX
DR   WPI; 2002-122074/16.
XX
PT   Human leukocyte antigen (HLA) typing, useful for judging HLA genotypes
PT   of individuals e.g. by determining immunogenetic differences when
PT   transplanting between them -
XX
PS   Claim 10; Page 146; 345pp; Japanese.
XX
CC   The invention relates to a typing kit for judging human leukocyte antigen
CC   (HLA) genotype of a sample by hybridising a substrate on which 10-24 base
CC   oligonucleotides (ABL30512-ABL31809) originating in the sequences of
CC   genes e.g. belonging to HLA class I antigens on human genome and
CC   containing gene polymorphisms as alloantigens have been immobilised as
CC   primers for amplification of cleaved nucleic acids relating to gene
CC   polymorphisms. The method is useful for judging HLA genotypes of
CC   individuals by determining immunogenetic differences before transplanting
CC   between them, providing genetic information to decide compatibility of
CC   organ and tissue for transplantation e.g. of bone marrow, kidney, liver,
CC   pancreas, Langerhans islet in pancreas and cornea, susceptibility
CC   diagnosis of genetic diseases and identifying individuals.
XX
SQ   Sequence 18 BP; 2 A; 2 C; 9 G; 5 T; 0 other;


  Query Match             100.0%;  Score 11;  DB 24;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||
Db      14 TCTCACCAACC 4


RESULT  13
AAQ06439/c
ID   AAQ06439 standard; DNA; 19 BP.
XX
AC   AAQ06439;
XX
DT   04-FEB-1991  (first entry)
XX
DE   Oligonucleotide probe to a lambda HLA-DR-5.0 sequence associated with
DE   human type I diabetes myelitis.
XX
KW   Insulin-dependent diabetes; systemic lupus erythematosus;
KW   Reiter's disease; ss.
XX
OS   Homo sapiens.
XX
```

PN US4965189-A.
XX
PD 23-OCT-1990.
XX
PF 01-JUL-1986; 86US-0880857.
XX
PR 01-JUL-1986; 86US-0880857.
XX
PA (UYMA-) UNIV MASSACHUESETTS.
XX
PI Owerbach D;
XX
DR WPI; 1990-341710/45.
XX
PT DQ beta gene oligo:nucleotide(s) - for detection of proclivity in
PT humans for development of type I diabetes mellitus
XX
PS Disclosure; Col 3; 17pp; English.
XX
CC Probe may be used in tests for proclivity towards autoimmune
CC diseases such as insulin dependent diabetes, Reiter's disease etc.
CC Probes are highly specific, even able to differentiate between
CC restriction fragments of identical size, and may also be used in
CC tissue typing.
XX
SQ Sequence 19 BP; 2 A; 3 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 11; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||
Db 14 TCTCACCAACC 4

RESULT 14
AAQ26219/c
ID AAQ26219 standard; DNA; 19 BP.
XX
AC AAQ26219;
XX
DT 04-JAN-1993 (first entry)
XX
DE HLA-DR beta sub-type tailed probe DRB115 hybridising region.
XX
KW Tissue typing; identity determination; disease susceptible; ss.
XX
OS Synthetic.
XX
PN WO9210589-A.
XX
PD 25-JUN-1992.
XX
PF 06-DEC-1991; 91WO-US09294.
XX
PR 06-DEC-1990; 90US-0623098.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
PI Apple RJ, Begovich AB, Bugawan T, Erlich HA, Griffith RL;
PI Scharf SJ;
XX
DR WPI; 1992-234644/28.
XX
PT Method for determining HLA-DR beta sub-type in DNA sample -
PT comprises amplification and hybridisation with probes and
PT primers, useful in tissue typing
XX
PS Example; Page 40; 90pp; English.
XX
CC The sequence is that of the hybridising region of tailed probe DRB115
CC for use in a method for determining HLA-DR beta sub-type in a nucleic
CC acid sample. The method allows specific nucleic acid sequences of the
CC second exon of HLA-DR beta genes to be amplified then probed for
CC identification of polymorphic sequences. The amplified DNA is useful for
CC typing homozgous or heterozygous samples from a variety of sources and
CC for detecting allelic variants not distinguishable by serological
CC methods. The typing system can be used in a reverse dot blot format which
CC is simple and rapid to perform, produces detectable signals in minutes
CC and can be utilised in tissue typing, determination of individual
CC identity and identifying disease susceptible individuals.
CC See also AAQ26092-Q26367.
XX
SQ Sequence 19 BP; 3 A; 2 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 13; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||
Db 15 TCTCACCAACC 5

RESULT 15
AAX79408/c
ID AAX79408 standard; DNA; 19 BP.
XX
AC AAX79408;
XX
DT 17-AUG-1999 (first entry)
XX
DE HLA-DR typing probe VAL86.
XX
KW Tissue typing; human leukocyte antigen; HLA; MHC; donor; allele; PCR;
KW major histocompatibility complex; bone marrow transplant; primer;
KW amplification; polymerase chain reaction; probe; polymorphism;
KW sequence-specific oligonucleotide probe hybridisation; ss.
XX
OS Synthetic.
XX
PN US5468611-A.
XX
PD 21-NOV-1995.
XX
PF 08-APR-1993; 93US-0045530.
XX
PR 27-JUN-1990; 90US-0544218.
PR 08-APR-1993; 93US-0045530.
XX
PA (BLOO-) BLOOD CENT RES FOUND INC.
XX
PI Baxter-Lowe LA, Gorski JA;
XX
DR WPI; 1996-010091/01.
XX
PT Improved method for HLA typing - by DNA amplification and
PT sequence-specific oligo:nucleotide hybridisation, used to select
PT bone marrow donors
XX
PS Disclosure; Column 19-20; 20pp; English.
XX
CC A novel method of typing the human leukocyte antigen (HLA) of the major
CC histocompatibility complex (MHC), esp. for typing donors for bone marrow
CC transplants, involves determining if the donor tissue HLA-DR alleles are
CC selected from the gp.: HLA-DRW52C, DR12a,b, DR3a,n, DR5a-e, DRNew1,
CC DR6a, DR8a-d, DRW53a-c, DR4a-f, DR7, DR9, DR2a-c B3, DR2a-d B1, DR10 and
CC DR1a-c. The method uses PCR to amplify these regions followed by
CC sequence-specific oligonucleotide probe hybridisation (SSOPH) using the
CC probes AAX79365-X79429. SSOPH allows detection of polymorphisms that
CC predict differences at a single amino acid level thus reducing errors
CC and improving the chance of successfully matching tissues.
XX

SQ Sequence 19 BP; 2 A; 3 C; 9 G; 5 T; 0 other;

Query Match 100.0%; Score 11; DB 17; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 tctcaccaacc 11
           ||||||||||| 
Db      14 TCTCACCAACC 4
```

Search completed: July 30, 2002, 00:01:20
Job time: 4931 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:22:44 ; Search time 2542.47 Seconds
(without alignments)
58.395 Million cell updates/sec

Title: US-09-530-663B-17
Perfect score: 11
Sequence: 1 tctcaccaacc 11

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	11	100.0	36	12	AZ493433	AZ493433 1M0328A12
c 2	11	100.0	98	10	F30845	F30845 HSPD21538 H
3	11	100.0	100	9	AI292201	AI292201 qm86e09.x
4	11	100.0	102	12	AZ233895	AZ233895 RPCI-23-8
c 5	11	100.0	108	9	AA184061	AA184061 mo96f04.r
c 6	11	100.0	109	9	AA258662	AA258662 zr61h08.r
7	11	100.0	111	10	BG953134	BG953134 CM4-CT063
c 8	11	100.0	120	10	BF828693	BF828693 MR2-HN003
c 9	11	100.0	122	10	BF085206	BF085206 MR3-GN002
c 10	11	100.0	123	9	AA325418	AA325418 EST28422
c 11	11	100.0	124	9	AA648082	AA648082 ns10b07.r
c 12	11	100.0	124	10	BI040954	BI040954 CM3-NT026
13	11	100.0	127	12	AZ753347	AZ753347 RPCI-24-1
14	11	100.0	129	12	AZ738998	AZ738998 RPCI-24-7
c 15	11	100.0	131	9	AI974770	AI974770 T113239e
c 16	11	100.0	131	10	BI301600	BI301600 UI-R-DL0-
c 17	11	100.0	135	10	BI562481	BI562481 603256389
c 18	11	100.0	136	9	AA052296	AA052296 mb92g02.r
c 19	11	100.0	136	10	BE807681	BE807681 ss29a01.y
20	11	100.0	137	10	BG342269	BG342269 602374360
c 21	11	100.0	139	10	BI052985	BI052985 RC0-GN027
22	11	100.0	145	9	BE033228	BE033228 133585 MA
c 23	11	100.0	145	10	BE980472	BE980472 UI-M-BG2-
24	11	100.0	146	9	AV028552	AV028552 AV028552
c 25	11	100.0	146	12	AQ997032	AQ997032 RPCI-23-3
c 26	11	100.0	148	10	BG954331	BG954331 CM4-CT065
c 27	11	100.0	151	9	AA061917	AA061917 mj84a04.r
c 28	11	100.0	152	10	BE715956	BE715956 MR2-HT075
29	11	100.0	154	10	BG133570	BG133570 EST466558
30	11	100.0	155	9	AI137789	AI137789 UI-R-E1-g
c 31	11	100.0	156	9	AV031373	AV031373 AV031373
32	11	100.0	156	9	AV237319	AV237319 AV237319
33	11	100.0	159	10	BF987685	BF987685 QV0-GN014
c 34	11	100.0	159	10	BM052472	BM052472 ic99b01.y
35	11	100.0	161	12	AZ447050	AZ447050 1M0243D14
36	11	100.0	161	12	B07010	B07010 cSRL-8h5-u
37	11	100.0	162	9	BE231730	BE231730 136547 MA
38	11	100.0	166	9	AA105067	AA105067 mm68b10.r
c 39	11	100.0	167	10	BG190599	BG190599 RST9672 A
c 40	11	100.0	170	9	AA576883	AA576883 nm78a10.s
41	11	100.0	173	10	BF645790	BF645790 NF028F04E
c 42	11	100.0	174	9	AA718915	AA718915 ah45a11.s
c 43	11	100.0	174	9	AI705839	AI705839 UI-R-AC1-
c 44	11	100.0	174	9	AW836760	AW836760 QV1-LT003
45	11	100.0	174	9	BE031039	BE031039 129375 MA

ALIGNMENTS

RESULT 1
AZ493433/c
LOCUS AZ493433 36 bp DNA linear GSS 05-OCT-2000
DEFINITION 1M0328A12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0328A12 F, DNA sequence.
ACCESSION AZ493433
VERSION AZ493433.1 GI:10667114
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 36)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0328 row: A column: 12
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 36.
FEATURES Location/Qualifiers
source 1. .36
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0328A12"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 6 a 4 c 16 g 10 t
ORIGIN

Query Match 100.0%; Score 11; DB 12; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||||
Db 29 TCTCACCAACC 19

RESULT 2
F30845/c
LOCUS F30845 98 bp mRNA linear EST 13-MAY-1999
DEFINITION HSPD21538 HM3 Homo sapiens cDNA clone s4000090D07, mRNA sequence.
ACCESSION F30845
VERSION F30845.1 GI:4816471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 98)
AUTHORS Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A., Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
TITLE Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT Contact: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at http://grup.bio.unipd.it.
FEATURES Location/Qualifiers
source 1. .98
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000090D07"
/clone_lib="HM3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/note="Vector: pcDNAII (Invitrogen); Site_1: BstXI; Site_2: NotI; The library was constructed by G. Lanfranchi. This library is not subtracted nor normalized. The first strand cDNA was primed with a biotinylated oligo-dT-NotI primer (5'-biotin-AACCCGGCTCGAGCGGCCGCTTTTTTTTTTTTTTTT-3'). The ds cDNA was sonicated and size-selected in the range 350-550 bp. The 3' specific fragments were selected by streptavidin coated magnetic beads, ligated to non-palindromic BstXI adapters, NotI digested and directionally cloned into BstXI-NotI cut pcDNAII vector."
BASE COUNT 42 a 11 c 21 g 24 t
ORIGIN

Query Match 100.0%; Score 11; DB 10; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||||
Db 67 TCTCACCAACC 57

RESULT 3
AI292201
LOCUS AI292201 100 bp mRNA linear EST 30-NOV-1998
DEFINITION qm86e09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1895656 3' similar to SW:TCPZ_HUMAN P40227 T-COMPLEX PROTEIN 1, ZETA SUBUNIT ;, mRNA sequence.
ACCESSION AI292201
VERSION AI292201.1 GI:3934975
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1895656"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 30 a 27 c 19 g 24 t
ORIGIN

Query Match 100.0%; Score 11; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11

```
          ||||||||||
Db     31 TCTCACCAACC 41


RESULT   4
AZ233895
LOCUS       AZ233895                 102 bp    DNA     linear   GSS 14-JUN-2000
DEFINITION  RPCI-23-82M10.TV RPCI-23 Mus musculus genomic clone RPCI-23-82M10,
            DNA sequence.
ACCESSION   AZ233895
VERSION     AZ233895.1  GI:8541941
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 102)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-82M10.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 82  row: M  column: 10
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .102
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-82M10"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       37 a     18 c     17 g     30 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 12;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     63 TCTCACCAACC 73


RESULT   5
AA184061/c
LOCUS       AA184061                 108 bp    mRNA    linear   EST 15-FEB-1997
DEFINITION  mo96f04.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:567583 5', mRNA sequence.
ACCESSION   AA184061
VERSION     AA184061.1  GI:1767429
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 108)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:342231
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 96.
FEATURES             Location/Qualifiers
     source          1. .108
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:567583"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       28 a     24 c     34 g     22 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 9;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     41 TCTCACCAACC 31


RESULT   6
AA258662/c
LOCUS       AA258662                 109 bp    mRNA    linear   EST 06-AUG-1997
DEFINITION  zr61h08.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:667935
            5', mRNA sequence.
ACCESSION   AA258662
VERSION     AA258662.1  GI:1893786
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
```

```
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1750   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .109
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:667935"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       27 a     23 c     31 g     28 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 9;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      20 TCTCACCAACC 10


RESULT   7
BG953134
LOCUS       BG953134                111 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  CM4-CT0639-220101-695-d02 CT0639 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG953134
VERSION     BG953134.1  GI:14371305
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 111)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-CT0639-
            220101-695-d02&t3=2001-01-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 111.
FEATURES             Location/Qualifiers
     source          1. .111
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0639"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       18 a     32 c     25 g     36 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 10;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      70 TCTCACCAACC 80


RESULT   8
BF828693/c
LOCUS       BF828693                120 bp    mRNA    linear   EST 13-JAN-2001
DEFINITION  MR2-HN0035-141200-015-e01 HN0035 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF828693
VERSION     BF828693.1  GI:12173470
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-HN0035-
            141200-015-e01&t3=2000-12-14&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 119.
FEATURES             Location/Qualifiers
     source          1. .120
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HN0035"
                     /dev_stage="Adult"
                     /note="Organ: head_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       28 a     31 c     38 g     22 t      1 others
ORIGIN


  Query Match             100.0%;  Score 11;  DB 10;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     28 TCTCACCAACC 18


RESULT  9
BF085206/c
LOCUS       BF085206                 122 bp    mRNA    linear   EST 19-OCT-2000
DEFINITION  MR3-GN0029-110900-004-f02_1 GN0029 Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BF085206
VERSION     BF085206.1  GI:10891007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 122)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR3-GN0029-110
            900-004-f02_1&t3=2000-09-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 54.
FEATURES             Location/Qualifiers
     source          1. .122
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0029"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT       48 a     21 c     32 g     21 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 10;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     39 TCTCACCAACC 29


RESULT  10
AA325418/c
LOCUS       AA325418                 123 bp    mRNA    linear   EST 20-APR-1997
DEFINITION  EST28422 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA325418
VERSION     AA325418.1  GI:1977683
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC168430
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .123
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):125942"
                     /db_xref="taxon:9606"
                     /clone_lib="Cerebellum II"
                     /tissue_type="cerebellum"
                     /dev_stage="adult"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       16 a     34 c     44 g     26 t      3 others
ORIGIN
```

```
  Query Match           100.0%;  Score 11;  DB 9;  Length 123;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     61 TCTCACCAACC 51


RESULT  11
AA648082/c
LOCUS       AA648082                 124 bp    mRNA    linear   EST 29-OCT-1997
DEFINITION  ns10b07.r1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:1183189
            similar to TR:G624778 G624778 E25. ;, mRNA sequence.
ACCESSION   AA648082
VERSION     AA648082.1  GI:2574511
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
            Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .124
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1183189"
                     /clone_lib="NCI_CGAP_Ew1"
                     /tissue_type="Ewing's sarcoma"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
                     cDNA made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert  size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       32 a     18 c     45 g     29 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 9;  Length 124;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     62 TCTCACCAACC 52


RESULT  12
BI040954/c
LOCUS       BI040954                 124 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  CM3-NT0268-130201-748-d05_1 NT0268 Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BI040954
VERSION     BI040954.1  GI:14447580
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-NT0268-
            130201-748-d05_1&t3=2001-02-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 119.
FEATURES             Location/Qualifiers
     source          1. .124
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0268"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       32 a     23 c     32 g     37 t
ORIGIN


  Query Match           100.0%;  Score 11;  DB 10;  Length 124;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     88 TCTCACCAACC 78


RESULT  13
AZ753347
LOCUS       AZ753347                 127 bp    DNA     linear   GSS 25-JAN-2001
DEFINITION  RPCI-24-147P15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-147P15
            , DNA sequence.
ACCESSION   AZ753347
VERSION     AZ753347.1  GI:12538506
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 127)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-147P15.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 147  row: P  column: 15
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .127
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-24-147P15"
                      /clone_lib="RPCI-24"
                      /sex="Male"
                      /cell_type="Spleen/Brain"
                      /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                      library was  cloned in the pTARBAC1 cloning vector at the
                      BamH1 sites using MboI  partially digested male C57BL/6J
                      DNA."
BASE COUNT       42 a     42 c     24 g     19 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 12;  Length 127;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      59 TCTCACCAACC 69


RESULT  14
AZ738998
LOCUS       AZ738998                 129 bp    DNA     linear   GSS 25-JAN-2001
DEFINITION  RPCI-24-71F6.TV RPCI-24 Mus musculus genomic clone RPCI-24-71F6,
            DNA sequence.
ACCESSION   AZ738998
VERSION     AZ738998.1  GI:12508671
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 129)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-71F6.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 71  row: F  column: 6
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .129
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-24-71F6"
                      /clone_lib="RPCI-24"
                      /sex="Male"
                      /cell_type="Spleen/Brain"
                      /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                      library was  cloned in the pTARBAC1 cloning vector at the
                      BamH1 sites using MboI  partially digested male C57BL/6J
                      DNA."
BASE COUNT       25 a     19 c     21 g     64 t
ORIGIN


  Query Match             100.0%;  Score 11;  DB 12;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||| |
Db      89 TCTCACCAACC 99


RESULT  15
AI974770/c
LOCUS       AI974770                 131 bp    mRNA    linear   EST 26-AUG-1999
DEFINITION  T113239e KV2 Medicago truncatula cDNA clone pKV2-1K8, mRNA
            sequence.
ACCESSION   AI974770
VERSION     AI974770.1  GI:5777151
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 131)
  AUTHORS   VandenBosch,K., Hur,J., Moore,J., Beremand,P., Peng,H. and Ellis,L.
  TITLE     ESTs from Sinorhizobium-inoculated roots of Medicago truncatula
            (1999b)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: VandenBosch K
            Department of Biology
            Texas A&M University
            College Station, TX 77843-3258, USA
            Tel: 409 845 7707
            Fax: 409 845 2891
            Email: kate@mail.bio.tamu.edu
            Other name: 02-KV2-3F4; date: 8/5/99; Submitted to the Database of
            Expressed Sequence Tags (dbEST) on 08/25/99; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .131
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pKV2-1K8"
                      /clone_lib="KV2"
                      /tissue_type="Seedling roots"
                      /dev_stage="2 days post-inoculation with Sinorhizobium
                      meliloti"
                      /lab_host="E. coli strain SOLR"
```

/note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unizap XR vector
from Stratagene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-Zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT 22 a 13 c 43 g 53 t
ORIGIN

Query Match 100.0%; Score 11; DB 9; Length 131;
Best Local Similarity 100.0%; Pred. No. 2.2e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctcaccaacc 11
||||||||||||
Db 112 TCTCACCAACC 102

Search completed: July 29, 2002, 23:22:49
Job time: 6825 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:56:14 ; Search time 65.09 Seconds
(without alignments)
41.511 Million cell updates/sec

Title: US-09-530-663B-17
Perfect score: 11
Sequence: 1 tctcaccaacc 11

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	11	100.0	12	1	US-08-025-038-16	Sequence 16, Appl
c	2	11	100.0	16	1	US-08-050-073-77	Sequence 77, Appl
c	3	11	100.0	16	1	US-08-050-073-168	Sequence 168, App
c	4	11	100.0	17	1	US-08-050-073-223	Sequence 223, App
c	5	11	100.0	19	1	US-08-050-073-191	Sequence 191, App
c	6	11	100.0	25	4	US-09-441-346A-10	Sequence 10, Appl
c	7	11	100.0	54	1	US-08-758-306-1354	Sequence 1354, Ap
c	8	11	100.0	69	3	US-08-463-903-76	Sequence 76, Appl
c	9	11	100.0	69	4	US-07-935-695-76	Sequence 76, Appl
	10	11	100.0	74	3	US-08-463-903-77	Sequence 77, Appl
	11	11	100.0	74	4	US-07-935-695-77	Sequence 77, Appl
	12	11	100.0	220	2	US-08-933-616-2	Sequence 2, Appli
c	13	11	100.0	268	1	US-08-039-137-16	Sequence 16, Appl
c	14	11	100.0	269	1	US-08-025-038-32	Sequence 32, Appl
c	15	11	100.0	269	1	US-08-025-038-33	Sequence 33, Appl
c	16	11	100.0	269	1	US-08-025-038-34	Sequence 34, Appl
c	17	11	100.0	269	1	US-08-025-038-37	Sequence 37, Appl
c	18	11	100.0	269	1	US-08-050-073-1	Sequence 1, Appli
c	19	11	100.0	269	1	US-08-050-073-3	Sequence 3, Appli
c	20	11	100.0	269	1	US-08-050-073-5	Sequence 5, Appli
c	21	11	100.0	269	1	US-08-050-073-7	Sequence 7, Appli
c	22	11	100.0	269	1	US-08-050-073-11	Sequence 11, Appl
c	23	11	100.0	269	1	US-08-050-073-13	Sequence 13, Appl
c	24	11	100.0	269	1	US-08-050-073-14	Sequence 14, Appl
c	25	11	100.0	269	1	US-08-050-073-15	Sequence 15, Appl
c	26	11	100.0	269	1	US-08-050-073-18	Sequence 18, Appl
c	27	11	100.0	269	1	US-08-050-073-19	Sequence 19, Appl
c	28	11	100.0	269	1	US-08-050-073-20	Sequence 20, Appl
c	29	11	100.0	269	1	US-08-050-073-21	Sequence 21, Appl
c	30	11	100.0	269	1	US-08-050-073-23	Sequence 23, Appl
c	31	11	100.0	269	1	US-08-050-073-24	Sequence 24, Appl
c	32	11	100.0	269	1	US-08-050-073-25	Sequence 25, Appl
c	33	11	100.0	269	1	US-08-050-073-29	Sequence 29, Appl
c	34	11	100.0	269	1	US-08-050-073-33	Sequence 33, Appl
c	35	11	100.0	269	1	US-08-050-073-34	Sequence 34, Appl
c	36	11	100.0	269	1	US-08-050-073-36	Sequence 36, Appl
c	37	11	100.0	269	1	US-08-050-073-38	Sequence 38, Appl
c	38	11	100.0	269	1	US-08-050-073-39	Sequence 39, Appl
c	39	11	100.0	269	1	US-08-050-073-43	Sequence 43, Appl
c	40	11	100.0	269	1	US-08-050-073-45	Sequence 45, Appl
c	41	11	100.0	269	1	US-08-050-073-46	Sequence 46, Appl
c	42	11	100.0	269	1	US-08-050-073-48	Sequence 48, Appl
c	43	11	100.0	269	1	US-08-050-073-52	Sequence 52, Appl
c	44	11	100.0	269	1	US-08-050-073-53	Sequence 53, Appl
c	45	11	100.0	269	1	US-08-050-073-69	Sequence 69, Appl

ALIGNMENTS

```
RESULT   1
US-08-025-038-16
; Sequence 16, Application US/08025038
; Patent No. 5545526
;  GENERAL INFORMATION:
;    APPLICANT:  BAXTER-LOWE, Lee-Ann
;    TITLE OF INVENTION:  Method For HLA Typing
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  777 E. Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-5367
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/025,038
;      FILING DATE:  19930301
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/544,218
;      FILING DATE:  27-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Philip G.
;      REGISTRATION NUMBER:  30,478
;      REFERENCE/DOCKET NUMBER:  204 854
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414)289-3761
;      TELEFAX:  (414)289-3791
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-025-038-16


  Query Match           100.0%;  Score 11;  DB 1;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tctcaccaacc 11
           ||||||||||||
Db       2 TCTCACCAACC 12
```

```
RESULT   2
US-08-050-073-77/c
; Sequence 77, Application US/08050073
; Patent No. 5567809
;  GENERAL INFORMATION:
;    APPLICANT:  Apple, Raymond J.
;    APPLICANT:  Begovich, Ann B.
;    APPLICANT:  Bugawan, Teodorica L.
;    APPLICANT:  Erlich, Henry A.
;    APPLICANT:  Griffith, Robert L.
;    APPLICANT:  Scharf, Stephen J.
;    TITLE OF INVENTION:  Methods and Reagents for HLA DRBeta DNA
;    TITLE OF INVENTION:  Typing
;    NUMBER OF SEQUENCES:  315
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,073
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petry, Douglas A.
;      REGISTRATION NUMBER:  35,321
;      REFERENCE/DOCKET NUMBER:  8769
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2974
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  77:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-050-073-77


  Query Match            100.0%;  Score 11;  DB 1;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     13 TCTCACCAACC 3


RESULT   3
US-08-050-073-168/c
; Sequence 168, Application US/08050073
; Patent No. 5567809
;  GENERAL INFORMATION:
;    APPLICANT:  Apple, Raymond J.
;    APPLICANT:  Begovich, Ann B.
;    APPLICANT:  Bugawan, Teodorica L.
;    APPLICANT:  Erlich, Henry A.
;    APPLICANT:  Griffith, Robert L.
;    APPLICANT:  Scharf, Stephen J.
;    TITLE OF INVENTION:  Methods and Reagents for HLA DRBeta DNA
;    TITLE OF INVENTION:  Typing
;    NUMBER OF SEQUENCES:  315
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,073
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petry, Douglas A.
;      REGISTRATION NUMBER:  35,321
;      REFERENCE/DOCKET NUMBER:  8769
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2974
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  168:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-050-073-168


  Query Match            100.0%;  Score 11;  DB 1;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db     13 TCTCACCAACC 3


RESULT   4
US-08-050-073-223/c
; Sequence 223, Application US/08050073
; Patent No. 5567809
;  GENERAL INFORMATION:
;    APPLICANT:  Apple, Raymond J.
;    APPLICANT:  Begovich, Ann B.
;    APPLICANT:  Bugawan, Teodorica L.
;    APPLICANT:  Erlich, Henry A.
;    APPLICANT:  Griffith, Robert L.
;    APPLICANT:  Scharf, Stephen J.
;    TITLE OF INVENTION:  Methods and Reagents for HLA DRBeta DNA
;    TITLE OF INVENTION:  Typing
;    NUMBER OF SEQUENCES:  315
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,073
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petry, Douglas A.
```

```
;      REGISTRATION NUMBER:  35,321
;      REFERENCE/DOCKET NUMBER:  8769
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2974
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  223:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-050-073-223


  Query Match           100.0%;  Score 11;  DB 1;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     14 TCTCACCAACC 4


RESULT   5
US-08-050-073-191/c
; Sequence 191, Application US/08050073
; Patent No. 5567809
;  GENERAL INFORMATION:
;    APPLICANT:  Apple, Raymond J.
;    APPLICANT:  Begovich, Ann B.
;    APPLICANT:  Bugawan, Teodorica L.
;    APPLICANT:  Erlich, Henry A.
;    APPLICANT:  Griffith, Robert L.
;    APPLICANT:  Scharf, Stephen J.
;    TITLE OF INVENTION:  Methods and Reagents for HLA DRBeta DNA
;    TITLE OF INVENTION:  Typing
;    NUMBER OF SEQUENCES:  315
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,073
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petry, Douglas A.
;      REGISTRATION NUMBER:  35,321
;      REFERENCE/DOCKET NUMBER:  8769
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2974
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  191:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-050-073-191


  Query Match           100.0%;  Score 11;  DB 1;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     15 TCTCACCAACC 5


RESULT   6
US-09-441-346A-10/c
; Sequence 10, Application US/09441346A
; Patent No. 6242588
; GENERAL INFORMATION:
;  APPLICANT:  Sheppard, Paul O.
;  APPLICANT:  Piddington, Christopher S.
;  APPLICANT:  Ellsworth, Jeff L.
;  TITLE OF INVENTION:  TESTIS-SPECIFIC GLYCOPROTEIN ZPEP10
;  FILE REFERENCE:  98-34
;  CURRENT APPLICATION NUMBER: US/09/441,346A
;  CURRENT FILING DATE:  1999-11-16
;  PRIOR APPLICATION NUMBER:  60/109,216
;  PRIOR FILING DATE:  1998-11-20
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide ZC16,187
US-09-441-346A-10


  Query Match           100.0%;  Score 11;  DB 4;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     22 TCTCACCAACC 12


RESULT   7
US-08-758-306-1354/c
; Sequence 1354, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  212/132
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1354:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  54 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-758-306-1354


  Query Match           100.0%;  Score 11;  DB 1;  Length 54;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     11 TCTCACCAACC 1


RESULT   8
US-08-463-903-76/c
; Sequence 76, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
;  APPLICANT: Mezes, Peter S.
;  APPLICANT: Richard, Ruth A.
;  APPLICANT: Affholter, Joseph A.
;  APPLICANT: Kotite, Nicolas J.
;  TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE: 40224A US
;  CURRENT APPLICATION NUMBER: US/08/463,903
;  CURRENT FILING DATE: 1995-06-05
;  EARLIER APPLICATION NUMBER: US 07/935,695
;  EARLIER FILING DATE: 1992-08-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 76
;   LENGTH: 69
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: 0101F primer
;   LOCATION: 1..69
US-08-463-903-76


  Query Match           100.0%;  Score 11;  DB 3;  Length 69;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     41 TCTCACCAACC 31


RESULT   9
US-07-935-695-76/c
; Sequence 76, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
;  APPLICANT:  Mezes, Peter S.
;  APPLICANT:  Richard, Ruth A.
;  APPLICANT:  Affholter, Joseph A.
;  APPLICANT:  Kotite, Nicolas J.
;  TITLE OF INVENTION:  Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE:  40224A US
;  CURRENT APPLICATION NUMBER: US/07/935,695
;  CURRENT FILING DATE:  1992-08-21
;  PRIOR APPLICATION NUMBER:  US 08/463,903
;  PRIOR FILING DATE:  1995-06-05
;  NUMBER OF SEQ ID NOS:  102
;  SOFTWARE:  MS-Word for Windows, Ver. 7.0
; SEQ ID NO 76
;   LENGTH: 69
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: 0101F primer
;   LOCATION: 1..69
;   OTHER INFORMATION: :
US-07-935-695-76


  Query Match           100.0%;  Score 11;  DB 4;  Length 69;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     41 TCTCACCAACC 31


RESULT  10
US-08-463-903-77
; Sequence 77, Application US/08463903
; Patent No. 6071515
; GENERAL INFORMATION:
;  APPLICANT: Mezes, Peter S.
;  APPLICANT: Richard, Ruth A.
;  APPLICANT: Affholter, Joseph A.
;  APPLICANT: Kotite, Nicolas J.
;  TITLE OF INVENTION: Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE: 40224A US
;  CURRENT APPLICATION NUMBER: US/08/463,903
;  CURRENT FILING DATE: 1995-06-05
;  EARLIER APPLICATION NUMBER: US 07/935,695
;  EARLIER FILING DATE: 1992-08-21
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: MS-Word for Windows, Ver. 7.0
; SEQ ID NO 77
;   LENGTH: 74
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: 0101R primer
;   LOCATION: 1..74
US-08-463-903-77


  Query Match           100.0%;  Score 11;  DB 3;  Length 74;
  Best Local Similarity  100.0%;  Pred. No. 3e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| |
Db     32 tctcaccaacc 42


RESULT  11
US-07-935-695-77
; Sequence 77, Application US/07935695
; Patent No. 6329507
; GENERAL INFORMATION:
;  APPLICANT:  Mezes, Peter S.
;  APPLICANT:  Richard, Ruth A.
;  APPLICANT:  Affholter, Joseph A.
;  APPLICANT:  Kotite, Nicolas J.
```

```
;  TITLE OF INVENTION:  Dimer and Multimer Forms of Single Chain Polypeptides
;  FILE REFERENCE:  40224A US
;  CURRENT APPLICATION NUMBER: US/07/935,695
;  CURRENT FILING DATE:  1992-08-21
;  PRIOR APPLICATION NUMBER:  US 08/463,903
;  PRIOR FILING DATE:  1995-06-05
;  NUMBER OF SEQ ID NOS:  102
;  SOFTWARE:  MS-Word for Windows, Ver. 7.0
; SEQ ID NO 77
;   LENGTH: 74
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: 0101R primer
;   LOCATION: 1..74
;   OTHER INFORMATION: :
US-07-935-695-77


  Query Match             100.0%;  Score 11;  DB 4;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     32 tctcaccaacc 42


RESULT  12
US-08-933-616-2
; Sequence 2, Application US/08933616
; Patent No. 5869331
;  GENERAL INFORMATION:
;    APPLICANT:  Dornburg, Ralph C.
;    TITLE OF INVENTION:  Cell-Type Specific Gene Transfer Using
;    TITLE OF INVENTION:  Retroviral Vectors Containing Antibody-Envelope Fusion
;    TITLE OF INVENTION:  Proteins and Wild-Type Envelope Fusion Proteins
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard R. Muccino
;      STREET:  P.O. Box 1267
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  08551
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,616
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/205,980
;      FILING DATE:  04-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Muccino, Richard R.
;      REGISTRATION NUMBER:  32,538
;      REFERENCE/DOCKET NUMBER:  UMD1-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 466-3407
;      TELEFAX:  (609) 466-2760
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  220 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-933-616-2


  Query Match             100.0%;  Score 11;  DB 2;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db     93 TCTCACCAACC 103


RESULT  13
US-08-039-137-16/c
; Sequence 16, Application US/08039137
; Patent No. 5759771
;  GENERAL INFORMATION:
;    APPLICANT:  Tilanus J.G., Marcel
;    TITLE OF INVENTION:  Method of Determining a Genotype by
;    TITLE OF INVENTION:  Comparing the Nucleotide Sequence of Members of a Gene
; Patent No. 5759771
;    TITLE OF INVENTION:  Family and Kit Therefor
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/039,137
;      FILING DATE:  14-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  0550-0024.10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  DR1 cDNA, Fig. 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..267
US-08-039-137-16


  Query Match             100.0%;  Score 11;  DB 1;  Length 268;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||| 
Db    247 TCTCACCAACC 237


RESULT  14
US-08-025-038-32/c
; Sequence 32, Application US/08025038
```

```
; Patent No. 5545526
;  GENERAL INFORMATION:
;    APPLICANT:  BAXTER-LOWE, Lee-Ann
;    TITLE OF INVENTION:  Method For HLA Typing
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  777 E. Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-5367
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/025,038
;      FILING DATE:  19930301
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/544,218
;      FILING DATE:  27-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Philip G.
;      REGISTRATION NUMBER:  30,478
;      REFERENCE/DOCKET NUMBER:  204 854
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414)289-3761
;      TELEFAX:  (414)289-3791
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-025-038-32


  Query Match           100.0%;  Score 11;  DB 1;  Length 269;
  Best Local Similarity  100.0%;  Pred. No. 3.2e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db    249 TCTCACCAACC 239


RESULT  15
US-08-025-038-33/c
; Sequence 33, Application US/08025038
; Patent No. 5545526
;  GENERAL INFORMATION:
;    APPLICANT:  BAXTER-LOWE, Lee-Ann
;    TITLE OF INVENTION:  Method For HLA Typing
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  777 E. Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-5367
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/025,038
;      FILING DATE:  19930301
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/544,218
;      FILING DATE:  27-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Philip G.
;      REGISTRATION NUMBER:  30,478
;      REFERENCE/DOCKET NUMBER:  204 854
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414)289-3761
;      TELEFAX:  (414)289-3791
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  269 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-025-038-33


  Query Match           100.0%;  Score 11;  DB 1;  Length 269;
  Best Local Similarity  100.0%;  Pred. No. 3.2e+02;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tctcaccaacc 11
          ||||||||||
Db    249 TCTCACCAACC 239


Search completed: July 29, 2002, 23:56:14
Job time: 4755 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 00:01:18 ; Search time 285.14 Seconds
(without alignments)
36.128 Million cell updates/sec

Title: US-09-530-663B-16
Perfect score: 6
Sequence: 1 ccgttc 6

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	6	100.0	10	16	AAQ99824	Loblolly pine fusi
c	2	6	100.0	10	20	AAX59797	Primer OPC6 for fu
	3	6	100.0	10	20	AAX23376	HLA-A, HLA-B, HLA-
	4	6	100.0	10	21	AAC81828	Gerbera flavone sy
	5	6	100.0	10	21	AAC81843	Gerbera flavone sy
	6	6	100.0	10	21	AAZ50866	Primer AP11 to ide
c	7	6	100.0	10	21	AAZ77645	Human dendritic ce
	8	6	100.0	10	21	AAZ77674	Human dendritic ce
c	9	6	100.0	10	21	AAZ78756	Human dendritic ce
c	10	6	100.0	10	21	AAZ79193	Human dendritic ce
	11	6	100.0	10	21	AAZ79337	Human dendritic ce
	12	6	100.0	10	21	AAZ81469	Metastatic breast
c	13	6	100.0	10	21	AAZ82818	Metastatic breast
	14	6	100.0	10	21	AAZ83972	Metastatic breast
	15	6	100.0	10	21	AAZ84187	Metastatic breast
c	16	6	100.0	10	21	AAZ85105	Metastatic breast
	17	6	100.0	10	21	AAZ86191	Metastatic breast
c	18	6	100.0	10	22	AAH63480	Human ubiquitously
c	19	6	100.0	10	22	AAH63759	Human ubiquitously
	20	6	100.0	10	22	AAC91836	C. cibarius SNGT2-
	21	6	100.0	10	22	AAC93023	C. cibarius SNGT2-
	22	6	100.0	10	22	AAF33902	Yeast NORF gene SA
c	23	6	100.0	10	22	AAF34378	Yeast NORF gene SA
	24	6	100.0	10	22	AAF35080	Yeast NORF gene SA
c	25	6	100.0	10	22	AAF35094	Yeast NORF gene SA
	26	6	100.0	10	22	AAF35249	Yeast NORF gene SA
	27	6	100.0	10	22	AAF35594	Yeast NORF gene SA
c	28	6	100.0	10	22	AAF36495	Yeast NORF gene SA
	29	6	100.0	10	22	AAF36565	Yeast NORF gene SA
	30	6	100.0	10	22	AAF38153	Yeast NORF gene SA
c	31	6	100.0	10	22	AAF39516	Yeast NORF gene SA
c	32	6	100.0	10	22	AAF39708	Yeast NORF gene SA
	33	6	100.0	10	22	AAF40052	Yeast NORF gene SA
c	34	6	100.0	10	22	AAF40961	Yeast NORF gene SA
	35	6	100.0	10	22	AAF41006	Yeast NORF gene SA
c	36	6	100.0	10	22	AAF41763	Yeast NORF gene SA
c	37	6	100.0	10	22	AAF41764	Yeast NORF gene SA
c	38	6	100.0	10	22	AAF41765	Yeast NORF gene SA
c	39	6	100.0	10	22	AAF41766	Yeast NORF gene SA
c	40	6	100.0	10	22	AAF42678	Yeast NORF gene SA
	41	6	100.0	10	22	AAF42824	Yeast NORF gene SA
c	42	6	100.0	10	24	AAS99282	Human F12 gene all
	43	6	100.0	12	13	AAQ24014	Herpesvirus inhibi
	44	6	100.0	12	15	AAQ65647	Primer/adaptor use
	45	6	100.0	12	17	AAT13204	Primer J Bam 12 us

ALIGNMENTS

RESULT 1
AAQ99824/c
ID AAQ99824 standard; cDNA; 10 BP.
XX
AC AAQ99824;
XX
DT 06-MAR-1996 (first entry)
XX
DE Loblolly pine fusiform rust disease resistance marker OPC6 primer.
XX
KW Loblolly pine; Pinus taeda; fusiform rust disease; resistance marker;
KW Cronartium quercum f.sp. fusiforme; Cqf; RAPD genetic marker;
KW random amplified polymorphic DNA analysis; woody perennial plant;
KW family selection; pedigree; mapping; primer; ss.
XX
OS Synthetic.
XX
PN WO9519697-A1.
XX
PD 27-JUL-1995.
XX
PF 19-JAN-1995; 95WO-US00677.
XX
PR 21-JAN-1994; 94US-0184567.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Grattapaglia D, O'Malley DM, Sederoff RR;
XX
DR WPI; 1995-269212/35.
XX
PT Determn. of heritable oligogenic traits in woody plants by genomic

PT mapping of multiple markers in a two generation plant family - used
PT to select plants with desired characteristics for breeding.
XX
PS Example 5; Page 31; 103pp; English.
XX
CC RAPD analysis was used to study resistance to particular strains of
CC Cronartium quercuum f.sp. fusiforme (Cqf), the causative agent of
CC fusiform rust disease, in loblolly pine (Pinus taeda). A putative
CC heterozygous mother tree (clone 10-5) and two open pollinated
CC daughters (half-sib clones 152-231 and 152-257) were crossed to a
CC highly susceptible pollen parent. Progeny were challenged with
CC inoculum from various aeciospore lines. It was found that the
CC marker amplified by the 10-mer primer in AAQ99824 was predictive of
CC resistance to inoculation with single Aeciospore line 2-36 in
CC clone 152-231 progeny. These and other results showed that resistance
CC to fusiform rust disease in loblolly pine is under oligogenic
CC control which can be mapped using genetic markers, using only a
CC two-generation pedigree.
XX
SQ Sequence 10 BP; 3 A; 3 C; 3 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 16; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 6 CCGTTC 1

RESULT 2
AAX59797/c
ID AAX59797 standard; DNA; 10 BP.
XX
AC AAX59797;
XX
DT 28-JUL-1999 (first entry)
XX
DE Primer OPC6 for fusiform rust disease resistance marker.
XX
KW Genetic marker; genetic locus; resistance; fusiform rust disease;
KW tree family; Pinus; PCR primer; ss.
XX
OS Synthetic.
XX
PN US5908978-A.
XX
PD 01-JUN-1999.
XX
PF 18-OCT-1995; 95US-0545253.
XX
PR 18-OCT-1995; 95US-0545253.
PR 21-JAN-1994; 94US-0184567.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Amerson HV, Grattapaglia D, Kuhlman EG, O'Malley DM;
PI Sederoff RR, Wilcox P;
XX
DR WPI; 1999-347038/29.
XX
PT Identifying resistance to fusiform rust disease in trees of the
PT genus Pinus
XX
PS Example 5; Column 23; 69pp; English.
XX
CC The specification describes a method of identifying a genetic marker
CC associated with a genetic locus conferring resistance to fusiform
CC rust disease in a family of trees of the genus Pinus. The method
CC comprises obtaining a sexually mature Pinus parent tree exhibiting
CC resistance to fusiform rust disease, obtaining a plurality of progeny
CC trees of the parent by self or cross-pollinations, assessing multiple
CC progeny trees for a number of genetic markers, identifying genetic
CC markers segregating in a Mendelian ratio and showing linkage with other
CC genetic markers, measuring resistance to fusiform rust disease in
CC multiple progeny trees and correlating the presence of resistance to
CC fusiform rust disease with at least one marker identified in the
CC previous step. The method is useful for determining the genetic basis
CC of resistance to fusiform rust disease and for producing trees of the
CC Pinus genus that are resistant to the disease. The present primer was
CC used in the method of the invention to identify and amplify resistance
CC markers.
XX
SQ Sequence 10 BP; 3 A; 3 C; 3 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 6 CCGTTC 1

RESULT 3
AAX23376
ID AAX23376 standard; DNA; 10 BP.
XX
AC AAX23376;
XX
DT 17-JUN-1999 (first entry)
XX
DE HLA-A, HLA-B, HLA-C polymorphism-specific primer E3b1 DNA.
XX
KW Primer; polymorphism; HLA-A; HLA-B; HLA-C; detection; probe array;
KW hybridisation pattern; ss.
XX
OS Synthetic.
XX
PN NL1006733-C2.
XX
PD 09-FEB-1999.
XX
PF 07-AUG-1997; 97NL-1006733.
XX
PR 07-AUG-1997; 97NL-1006733.
XX
PA (LEES/) LEE S H.
XX
PI Lee SH;
XX
DR WPI; 1999-213125/18.
XX
PT Hybridisation assay for identifying alleles - using array of
PT polymorphism-specific oligo:nucleotide probes
XX
PS Claim 16; Page 26; 67pp; Dutch.
XX
CC This invention describes a method for identifying an HLA-A, HLA-B or
CC HLA-C allele by detecting polymorphisms using the primers
CC AAX23371-X23380. The method involves (a) preparing single-stranded
CC nucleic acid molecules corresponding in sequence to the portion of the
CC allele containing the target polymorphisms, (b) labelling the
CC single-stranded nucleic acid molecules, immobilising oligonucleotide
CC probes, each specific for a known polymorphism, on a support and (c)
CC detecting fully complementary duplexes formed between the labelled
CC single-stranded nucleic acid molecules and the immobilised probes. The
CC method is used for determining differences and correspondences in
CC polymorphisms between individuals, tissues or organs by comparing
CC hybridisation patterns produced by the above method. The method uses
CC probe arrays rather than requiring sequential hybridisation and removal
CC of individual labelled probes.
XX
SQ Sequence 10 BP; 0 A; 5 C; 1 G; 4 T; 0 other;

```
  Query Match             100.0%;  Score 6;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      5 ccgttc 10


RESULT   4
AAC81828
ID   AAC81828 standard; DNA; 10 BP.
XX
AC   AAC81828;
XX
DT   22-FEB-2001  (first entry)
XX
DE   Gerbera flavone synthase FNSII primer decamer 6.
XX
KW   Gerbera; transgenic plant; flavone synthase II; FNSII; anticancer;
KW   immunomodulator; naringenin; apigenin; ornamental plant; flower colour;
KW   pharmaceutical; cancer; treatment; primer; ss.
XX
OS   Gerbera hybrida.
XX
PN   DE19918365-A1.
XX
PD   26-OCT-2000.
XX
PF   22-APR-1999;   99DE-1018365.
XX
PR   22-APR-1999;   99DE-1018365.
XX
PA   (MART/) MARTENS S.
PA   (FORK/) FORKMANN G.
XX
PI   Martens S,  Forkmann G;
XX
DR   WPI; 2000-648348/63.
XX
PT   New nucleic acid encoding flavone synthase II, useful e.g. for
PT   producing transgenic plants with altered flower color or flavone
PT   content  -
XX
PS   Example 4; Page 22; 40pp; German.
XX
CC   This invention describes a novel nucleic acid (I) that encodes flavone
CC   synthase II (FNSII) which has anticancer and immunomodulatory activity.
CC   FNSII catalyses conversion of naringenin to apigenin. (I) is used to
CC   produce transgenic ornamental plants that have targeted alterations in
CC   flower color, also altered content/distribution of flavones in leaves,
CC   flowers and other tissues, e.g. increased resistance properties or
CC   symbiotic capacity. FNSII expressed by (I) is used in synthesis of
CC   flavones that are useful as pharmaceuticals, e.g. in cancer treatment, or
CC   as biologically active substances, e.g. to improve the immune defence
CC   system. Oligonucleotide fragments of (I) are used as probes and primers,
CC   or as antisense or ribozyme agents for regulating expression of (II).
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;


  Query Match             100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      3 ccgttc 8


RESULT   5
AAC81843
ID   AAC81843 standard; DNA; 10 BP.
XX
AC   AAC81843;
XX
DT   22-FEB-2001  (first entry)
XX
DE   Gerbera flavone synthase FNSII haem-binding DNA fragment D6.
XX
KW   Gerbera; transgenic plant; flavone synthase II; FNSII; anticancer;
KW   immunomodulator; naringenin; apigenin; ornamental plant; flower colour;
KW   pharmaceutical; cancer; treatment; ss.
XX
OS   Gerbera hybrida.
XX
PN   DE19918365-A1.
XX
PD   26-OCT-2000.
XX
PF   22-APR-1999;   99DE-1018365.
XX
PR   22-APR-1999;   99DE-1018365.
XX
PA   (MART/) MARTENS S.
PA   (FORK/) FORKMANN G.
XX
PI   Martens S,  Forkmann G;
XX
DR   WPI; 2000-648348/63.
XX
PT   New nucleic acid encoding flavone synthase II, useful e.g. for
PT   producing transgenic plants with altered flower color or flavone
PT   content  -
XX
PS   Disclosure; Fig 4B; 40pp; German.
XX
CC   This invention describes a novel nucleic acid (I) that encodes flavone
CC   synthase II (FNSII) which has anticancer and immunomodulatory activity.
CC   FNSII catalyses conversion of naringenin to apigenin. (I) is used to
CC   produce transgenic ornamental plants that have targeted alterations in
CC   flower color, also altered content/distribution of flavones in leaves,
CC   flowers and other tissues, e.g. increased resistance properties or
CC   symbiotic capacity. FNSII expressed by (I) is used in synthesis of
CC   flavones that are useful as pharmaceuticals, e.g. in cancer treatment, or
CC   as biologically active substances, e.g. to improve the immune defence
CC   system. Oligonucleotide fragments of (I) are used as probes and primers,
CC   or as antisense or ribozyme agents for regulating expression of (II).
XX
SQ   Sequence 10 BP; 0 A; 4 C; 4 G; 2 T; 0 other;


  Query Match             100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      3 ccgttc 8


RESULT   6
AAZ50866
ID   AAZ50866 standard; DNA; 10 BP.
XX
AC   AAZ50866;
XX
DT   31-MAY-2000  (first entry)
XX
DE   Primer AP11 to identify tobacco salicylic acid inducible genes.
XX
KW   Tobacco plant; salicylic acid inducible gene; fungal pathogen;
KW   SA-inducible gene; transgenic plant; pathogen resistance; PCR primer; ss.
XX
```

OS Nicotiana tabacum.
XX
PN WO200008186-A1.
XX
PD 17-FEB-2000.
XX
PF 02-AUG-1999; 99WO-EP05581.
XX
PR 03-AUG-1998; 98US-0095187.
XX
PA (MOGE-) MOGEN INT NV.
XX
PI Stuiver MH, Jepson I, Horvath DM, Chua N;
XX
DR WPI; 2000-205725/18.
XX
PT Novel salicylic acid inducible genes from tobacco plants, useful for
PT making transgenic plants with enhanced pathogenic resistance -
XX
PS Example 1; Page 53; 57pp; English.
XX
CC The patent discloses fifteen new salicylic acid (SA) inducible genes from
CC Nicotiana tabacum, nine of which were subcloned and sequenced. Based on
CC different kinetics of induction response, these genes were classified
CC into four categories, class I, II, III and IV response genes. The
CC SA-inducible genes are useful for making transgenic plants with enhanced
CC pathogen resistance. The plants incorporating these genes show reduced
CC susceptibility to fungal pathogens. The present sequence is an
CC upstream primer AP11 used in differential display PCR reactions along
CC with downstream primers T12MG or T12MC to identify tobacco
CC SA-inducible genes.
XX
SQ Sequence 10 BP; 2 A; 4 C; 2 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 5 ccgttc 10

RESULT 7
AAZ77645/c
ID AAZ77645 standard; DNA; 10 BP.
XX
AC AAZ77645;
XX
DT 10-APR-2000 (first entry)
XX
DE Human dendritic cell SAGE tag, SEQ ID NO:73.
XX
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW APC; monocyte-derived dendritic cell; differential gene expression;
KW immunostimulatory cofactor; costimulatory factor; CTL;
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS Homo sapiens.
XX
PN WO9965924-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13800.
XX
PR 19-JUN-1998; 98US-0089833.
PR 19-JUN-1998; 98US-0089844.
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089878.
PR 19-JUN-1998; 98US-0089991.
PR 19-JUN-1998; 98US-0089992.
PR 19-JUN-1998; 98US-0089993.
PR 19-JUN-1998; 98US-0089994.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0089999.
PR 19-JUN-1998; 98US-0090000.
PR 19-JUN-1998; 98US-0090035.
PR 19-JUN-1998; 98US-0090036.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
PR 19-JUN-1998; 98US-0090042.
PR 19-JUN-1998; 98US-0090043.
PR 19-JUN-1998; 98US-0090044.
PR 19-JUN-1998; 98US-0090045.
PR 19-JUN-1998; 98US-0090047.
PR 19-JUN-1998; 98US-0090048.
PR 19-JUN-1998; 98US-0090072.
PR 19-JUN-1998; 98US-0090076.
PR 19-JUN-1998; 98US-0090077.
PR 19-JUN-1998; 98US-0090078.
PR 19-JUN-1998; 98US-0090079.
PR 19-JUN-1998; 98US-0090080.
PR 08-DEC-1998; 98US-0111715.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106077/09.
XX
PT Isolated polynucleotides differentially expressed in antigen-presenting
PT cells, useful in gene vaccines against cancer -
XX
PS Claim 1; Page 65; 130pp; English.
XX
CC Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC expression) tags used to identify mRNA transcripts encoding
CC immunostimulatory cofactor proteins which are preferentially or
CC differentially expressed in monocyte-derived dendritic cells compared
CC with monocytes. Some of the transcripts correspond to known genes or
CC ESTs (expressed sequence tags) which were previously unknown to be
CC preferentially or differentially expressed in dendritic cells, while
CC other transcripts correspond to novel genes. Antigen-presenting cell
CC (APC)-associated costimulatory factors play an important role in the
CC activation of the cytotoxic immune response, particularly against tumour
CC cells. Tumour antigen presentation via the MHC (major histocompatibility
CC complex) and subsequent recognition by T-cell receptors is alone
CC insufficient to activate a robust cytotoxic immune response that can
CC lyse the tumour cells, immunostimulatory cofactors also being required
CC for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC sequences identified using the SAGE tags have several potential uses.
CC They may be used in vaccines to induce an immune response, particularly
CC against a tumour antigen; to modulate the genotype of an APC; to screen
CC for agents that modulate expression of differentially expressed genes in
CC an APC; and as hybridisation probes/amplification primers for the
CC diagnosis, prognosis and monitoring of diseases related to abnormal
CC expression of these genes. Detection of the dendritic cell
CC differentially expressed genes, or of their encoded proteins, can be used
CC to identify cells as belonging to the monocyte lineage. Cells containing
CC these genes can be used in active immunotherapy (or to stimulate
CC production of a population of antigen-specific effector cells) and
CC vectors containing them are used in gene therapy. Co-administration of
CC tumour antigens and APC-associated costimulatory factors ensures adequate
CC antigen presentation to endogenous APCs and upregulates the APCs for the
CC presentation of co-stimulatory signals, migration to T cell-rich sites,
CC secretion of T cell growth factors and secretion of chemokines for
CC recruitment of immune effector cells.
XX
SQ Sequence 10 BP; 3 A; 1 C; 6 G; 0 U; 0 other;

```
  Query Match             100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      8 CCGTTC 3


RESULT   8
AAZ77674
ID   AAZ77674 standard; DNA; 10 BP.
XX
AC   AAZ77674;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human dendritic cell SAGE tag, SEQ ID NO:102.
XX
KW   SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW   APC; monocyte-derived dendritic cell; differential gene expression;
KW   immunostimulatory cofactor; costimulatory factor; CTL;
KW   cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965924-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13800.
XX
PR   19-JUN-1998;   98US-0089833.
PR   19-JUN-1998;   98US-0089844.
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089878.
PR   19-JUN-1998;   98US-0089991.
PR   19-JUN-1998;   98US-0089992.
PR   19-JUN-1998;   98US-0089993.
PR   19-JUN-1998;   98US-0089994.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0089999.
PR   19-JUN-1998;   98US-0090000.
PR   19-JUN-1998;   98US-0090035.
PR   19-JUN-1998;   98US-0090036.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
PR   19-JUN-1998;   98US-0090042.
PR   19-JUN-1998;   98US-0090043.
PR   19-JUN-1998;   98US-0090044.
PR   19-JUN-1998;   98US-0090045.
PR   19-JUN-1998;   98US-0090047.
PR   19-JUN-1998;   98US-0090048.
PR   19-JUN-1998;   98US-0090072.
PR   19-JUN-1998;   98US-0090076.
PR   19-JUN-1998;   98US-0090077.
PR   19-JUN-1998;   98US-0090078.
PR   19-JUN-1998;   98US-0090079.
PR   19-JUN-1998;   98US-0090080.
PR   08-DEC-1998;   98US-0111715.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106077/09.
XX
PT   Isolated polynucleotides differentially expressed in antigen-presenting
PT   cells, useful in gene vaccines against cancer -
XX
PS   Claim 1; Page 66; 130pp; English.
XX
CC   Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC   expression) tags used to identify mRNA transcripts encoding
CC   immunostimulatory cofactor proteins which are preferentially or
CC   differentially expressed in monocyte-derived dendritic cells compared
CC   with monocytes. Some of the transcripts correspond to known genes or
CC   ESTs (expressed sequence tags) which were previously unknown to be
CC   preferentially or differentially expressed in dendritic cells, while
CC   other transcripts correspond to novel genes. Antigen-presenting cell
CC   (APC)-associated costimulatory factors play an important role in the
CC   activation of the cytotoxic immune response, particularly against tumour
CC   cells. Tumour antigen presentation via the MHC (major histocompatibility
CC   complex) and subsequent recognition by T-cell receptors is alone
CC   insufficient to activate a robust cytotoxic immune response that can
CC   lyse the tumour cells, immunostimulatory cofactors also being required
CC   for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC   sequences identified using the SAGE tags have several potential uses.
CC   They may be used in vaccines to induce an immune response, particularly
CC   against a tumour antigen; to modulate the genotype of an APC; to screen
CC   for agents that modulate expression of differentially expressed genes in
CC   an APC; and as hybridisation probes/amplification primers for the
CC   diagnosis, prognosis and monitoring of diseases related to abnormal
CC   expression of these genes. Detection of the dendritic cell
CC   differentially expressed genes, or of their encoded proteins, can be used
CC   to identify cells as belonging to the monocyte lineage. Cells containing
CC   these genes can be used in active immunotherapy (or to stimulate
CC   production of a population of antigen-specific effector cells) and
CC   vectors containing them are used in gene therapy. Co-administration of
CC   tumour antigens and APC-associated costimulatory factors ensures adequate
CC   antigen presentation to endogenous APCs and upregulates the APCs for the
CC   presentation of co-stimulatory signals, migration to T cell-rich sites,
CC   secretion of T cell growth factors and secretion of chemokines for
CC   recruitment of immune effector cells.
XX
SQ   Sequence 10 BP; 1 A; 3 C; 2 G; 4 T; 0 other;


  Query Match             100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      2 ccgttc 7


RESULT   9
AAZ78756/c
ID   AAZ78756 standard; DNA; 10 BP.
XX
AC   AAZ78756;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human dendritic cell SAGE tag, SEQ ID NO:1184.
XX
KW   SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW   APC; monocyte-derived dendritic cell; differential gene expression;
KW   immunostimulatory cofactor; costimulatory factor; CTL;
KW   cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965924-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13800.
XX
PR   19-JUN-1998;   98US-0089833.
PR   19-JUN-1998;   98US-0089844.
PR   19-JUN-1998;   98US-0089853.
```

PR 19-JUN-1998; 98US-0089878.
PR 19-JUN-1998; 98US-0089991.
PR 19-JUN-1998; 98US-0089992.
PR 19-JUN-1998; 98US-0089993.
PR 19-JUN-1998; 98US-0089994.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0089999.
PR 19-JUN-1998; 98US-0090000.
PR 19-JUN-1998; 98US-0090035.
PR 19-JUN-1998; 98US-0090036.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
PR 19-JUN-1998; 98US-0090042.
PR 19-JUN-1998; 98US-0090043.
PR 19-JUN-1998; 98US-0090044.
PR 19-JUN-1998; 98US-0090045.
PR 19-JUN-1998; 98US-0090047.
PR 19-JUN-1998; 98US-0090048.
PR 19-JUN-1998; 98US-0090072.
PR 19-JUN-1998; 98US-0090076.
PR 19-JUN-1998; 98US-0090077.
PR 19-JUN-1998; 98US-0090078.
PR 19-JUN-1998; 98US-0090079.
PR 19-JUN-1998; 98US-0090080.
PR 08-DEC-1998; 98US-0111715.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106077/09.
XX
PT Isolated polynucleotides differentially expressed in antigen-presenting
PT cells, useful in gene vaccines against cancer -
XX
PS Claim 1; Page 98; 130pp; English.
XX
CC Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC expression) tags used to identify mRNA transcripts encoding
CC immunostimulatory cofactor proteins which are preferentially or
CC differentially expressed in monocyte-derived dendritic cells compared
CC with monocytes. Some of the transcripts correspond to known genes or
CC ESTs (expressed sequence tags) which were previously unknown to be
CC preferentially or differentially expressed in dendritic cells, while
CC other transcripts correspond to novel genes. Antigen-presenting cell
CC (APC)-associated costimulatory factors play an important role in the
CC activation of the cytotoxic immune response, particularly against tumour
CC cells. Tumour antigen presentation via the MHC (major histocompatibility
CC complex) and subsequent recognition by T-cell receptors is alone
CC insufficient to activate a robust cytotoxic immune response that can
CC lyse the tumour cells, immunostimulatory cofactors also being required
CC for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC sequences identified using the SAGE tags have several potential uses.
CC They may be used in vaccines to induce an immune response, particularly
CC against a tumour antigen; to modulate the genotype of an APC; to screen
CC for agents that modulate expression of differentially expressed genes in
CC an APC; and as hybridisation probes/amplification primers for the
CC diagnosis, prognosis and monitoring of diseases related to abnormal
CC expression of these genes. Detection of the dendritic cell
CC differentially expressed genes, or of their encoded proteins, can be used
CC to identify cells as belonging to the monocyte lineage. Cells containing
CC these genes can be used in active immunotherapy (or to stimulate
CC production of a population of antigen-specific effector cells) and
CC vectors containing them are used in gene therapy. Co-administration of
CC tumour antigens and APC-associated costimulatory factors ensures adequate
CC antigen presentation to endogenous APCs and upregulates the APCs for the
CC presentation of co-stimulatory signals, migration to T cell-rich sites,
CC secretion of T cell growth factors and secretion of chemokines for
CC recruitment of immune effector cells.
XX

SQ Sequence 10 BP; 3 A; 2 C; 5 G; 0 U; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 9 CCGTTC 4

RESULT 10
AAZ79193/c
ID AAZ79193 standard; DNA; 10 BP.
XX
AC AAZ79193;
XX
DT 10-APR-2000 (first entry)
XX
DE Human dendritic cell SAGE tag, SEQ ID NO:1621.
XX
KW SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW APC; monocyte-derived dendritic cell; differential gene expression;
KW immunostimulatory cofactor; costimulatory factor; CTL;
KW cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS Homo sapiens.
XX
PN WO9965924-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13800.
XX
PR 19-JUN-1998; 98US-0089833.
PR 19-JUN-1998; 98US-0089844.
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089878.
PR 19-JUN-1998; 98US-0089991.
PR 19-JUN-1998; 98US-0089992.
PR 19-JUN-1998; 98US-0089993.
PR 19-JUN-1998; 98US-0089994.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0089999.
PR 19-JUN-1998; 98US-0090000.
PR 19-JUN-1998; 98US-0090035.
PR 19-JUN-1998; 98US-0090036.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
PR 19-JUN-1998; 98US-0090042.
PR 19-JUN-1998; 98US-0090043.
PR 19-JUN-1998; 98US-0090044.
PR 19-JUN-1998; 98US-0090045.
PR 19-JUN-1998; 98US-0090047.
PR 19-JUN-1998; 98US-0090048.
PR 19-JUN-1998; 98US-0090072.
PR 19-JUN-1998; 98US-0090076.
PR 19-JUN-1998; 98US-0090077.
PR 19-JUN-1998; 98US-0090078.
PR 19-JUN-1998; 98US-0090079.
PR 19-JUN-1998; 98US-0090080.
PR 08-DEC-1998; 98US-0111715.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106077/09.
XX

```
PT   Isolated polynucleotides differentially expressed in antigen-presenting
PT   cells, useful in gene vaccines against cancer -
XX
PS   Claim 1; Page 111; 130pp; English.
XX
CC   Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC   expression) tags used to identify mRNA transcripts encoding
CC   immunostimulatory cofactor proteins which are preferentially or
CC   differentially expressed in monocyte-derived dendritic cells compared
CC   with monocytes. Some of the transcripts correspond to known genes or
CC   ESTs (expressed sequence tags) which were previously unknown to be
CC   preferentially or differentially expressed in dendritic cells, while
CC   other transcripts correspond to novel genes. Antigen-presenting cell
CC   (APC)-associated costimulatory factors play an important role in the
CC   activation of the cytotoxic immune response, particularly against tumour
CC   cells. Tumour antigen presentation via the MHC (major histocompatibility
CC   complex) and subsequent recognition by T-cell receptors is alone
CC   insufficient to activate a robust cytotoxic immune response that can
CC   lyse the tumour cells, immunostimulatory cofactors also being required
CC   for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC   sequences identified using the SAGE tags have several potential uses.
CC   They may be used in vaccines to induce an immune response, particularly
CC   against a tumour antigen; to modulate the genotype of an APC; to screen
CC   for agents that modulate expression of differentially expressed genes in
CC   an APC; and as hybridisation probes/amplification primers for the
CC   diagnosis, prognosis and monitoring of diseases related to abnormal
CC   expression of these genes. Detection of the dendritic cell
CC   differentially expressed genes, or of their encoded proteins, can be used
CC   to identify cells as belonging to the monocyte lineage. Cells containing
CC   these genes can be used in active immunotherapy (or to stimulate
CC   production of a population of antigen-specific effector cells) and
CC   vectors containing them are used in gene therapy. Co-administration of
CC   tumour antigens and APC-associated costimulatory factors ensures adequate
CC   antigen presentation to endogenous APCs and upregulates the APCs for the
CC   presentation of co-stimulatory signals, migration to T cell-rich sites,
CC   secretion of T cell growth factors and secretion of chemokines for
CC   recruitment of immune effector cells.
XX
SQ   Sequence 10 BP; 3 A; 2 C; 4 G; 1 T; 0 other;


  Query Match             100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db      10 CCGTTC 5


RESULT  11
AAZ79337
ID   AAZ79337 standard; DNA; 10 BP.
XX
AC   AAZ79337;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human dendritic cell SAGE tag, SEQ ID NO:1765.
XX
KW   SAGE tag; serial analysis of gene expression; antigen-presenting cell;
KW   APC; monocyte-derived dendritic cell; differential gene expression;
KW   immunostimulatory cofactor; costimulatory factor; CTL;
KW   cytotoxic T-lymphocyte; tumour antigen; immunotherapy; anticancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965924-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13800.
XX
PR   19-JUN-1998;   98US-0089833.
PR   19-JUN-1998;   98US-0089844.
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089878.
PR   19-JUN-1998;   98US-0089991.
PR   19-JUN-1998;   98US-0089992.
PR   19-JUN-1998;   98US-0089993.
PR   19-JUN-1998;   98US-0089994.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0089999.
PR   19-JUN-1998;   98US-0090000.
PR   19-JUN-1998;   98US-0090035.
PR   19-JUN-1998;   98US-0090036.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
PR   19-JUN-1998;   98US-0090042.
PR   19-JUN-1998;   98US-0090043.
PR   19-JUN-1998;   98US-0090044.
PR   19-JUN-1998;   98US-0090045.
PR   19-JUN-1998;   98US-0090047.
PR   19-JUN-1998;   98US-0090048.
PR   19-JUN-1998;   98US-0090072.
PR   19-JUN-1998;   98US-0090076.
PR   19-JUN-1998;   98US-0090077.
PR   19-JUN-1998;   98US-0090078.
PR   19-JUN-1998;   98US-0090079.
PR   19-JUN-1998;   98US-0090080.
PR   08-DEC-1998;   98US-0111715.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106077/09.
XX
PT   Isolated polynucleotides differentially expressed in antigen-presenting
PT   cells, useful in gene vaccines against cancer -
XX
PS   Claim 1; Page 115; 130pp; English.
XX
CC   Sequences AAZ77573-Z79709 represent SAGE (serial analysis of gene
CC   expression) tags used to identify mRNA transcripts encoding
CC   immunostimulatory cofactor proteins which are preferentially or
CC   differentially expressed in monocyte-derived dendritic cells compared
CC   with monocytes. Some of the transcripts correspond to known genes or
CC   ESTs (expressed sequence tags) which were previously unknown to be
CC   preferentially or differentially expressed in dendritic cells, while
CC   other transcripts correspond to novel genes. Antigen-presenting cell
CC   (APC)-associated costimulatory factors play an important role in the
CC   activation of the cytotoxic immune response, particularly against tumour
CC   cells. Tumour antigen presentation via the MHC (major histocompatibility
CC   complex) and subsequent recognition by T-cell receptors is alone
CC   insufficient to activate a robust cytotoxic immune response that can
CC   lyse the tumour cells, immunostimulatory cofactors also being required
CC   for efficient activation of cytotoxic T-lymphocytes (CTLs). Nucleic acid
CC   sequences identified using the SAGE tags have several potential uses.
CC   They may be used in vaccines to induce an immune response, particularly
CC   against a tumour antigen; to modulate the genotype of an APC; to screen
CC   for agents that modulate expression of differentially expressed genes in
CC   an APC; and as hybridisation probes/amplification primers for the
CC   diagnosis, prognosis and monitoring of diseases related to abnormal
CC   expression of these genes. Detection of the dendritic cell
CC   differentially expressed genes, or of their encoded proteins, can be used
CC   to identify cells as belonging to the monocyte lineage. Cells containing
CC   these genes can be used in active immunotherapy (or to stimulate
CC   production of a population of antigen-specific effector cells) and
CC   vectors containing them are used in gene therapy. Co-administration of
CC   tumour antigens and APC-associated costimulatory factors ensures adequate
CC   antigen presentation to endogenous APCs and upregulates the APCs for the
CC   presentation of co-stimulatory signals, migration to T cell-rich sites,
```

```
CC   secretion of T cell growth factors and secretion of chemokines for
CC   recruitment of immune effector cells.
XX
SQ   Sequence 10 BP; 1 A; 3 C; 3 G; 3 T; 0 other;


  Query Match           100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      1 ccgttc 6


RESULT  12
AAZ81469
ID   AAZ81469 standard; DNA; 10 BP.
XX
AC   AAZ81469;
XX
DT   07-APR-2000  (first entry)
XX
DE   Metastatic breast tumour cell upregulated transcript tag #703.
XX
KW   Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW   non-metastatic breast tumour tissue; gene therapy; anticancer;
KW   antimetastatic; vaccine; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965928-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13647.
XX
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106079/09.
XX
PT   Isolated polynucleotides differentially expressed between metastatic
PT   and non-metastatic breast cancer cells, useful for diagnosis,
PT   prevention and treatment of cancer -
XX
PS   Claim 1; Page 77; 219pp; English.
XX
CC   AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC   transcripts that are preferentially transcribed in the metastatic breast
CC   tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC   AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC   that are preferentially transcribed in the primary or non-metastatic
CC   breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC   cells). These transcripts can be used for diagnosis, prognosis,
CC   monitoring and treatment of breast cancer, particularly where metastatic.
CC   Diagnosis is by standard immunoassays or hybridisation/amplification
CC   reactions. Compounds that modulate expression of the transcripts are
CC   potentially useful for treatment of (metastatic) breast cancer, while
CC   promoters from the transcripts are used to direct expression, in selected
CC   cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC   sequences), particularly an antigen-encoding sequence for use in gene or
CC   cell-based vaccines. Polypeptides encoded by the transcripts are also
CC   useful in vaccines; for diagnosing breast cancer and for raising
CC   specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC   therapeutic agents. Host cells that produce the polypeptides can be used
CC   to expand and isolate populations of educated, antigen-specific immune
CC   effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC   adoptive immunotherapy.
XX
SQ   Sequence 10 BP; 2 A; 4 C; 2 G; 2 T; 0 other;


  Query Match           100.0%;  Score 6;  DB 21;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      1 ccgttc 6


RESULT  13
AAZ82818/c
ID   AAZ82818 standard; DNA; 10 BP.
XX
AC   AAZ82818;
XX
DT   07-APR-2000  (first entry)
XX
DE   Metastatic breast tumour cell upregulated transcript tag #2052.
XX
KW   Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW   non-metastatic breast tumour tissue; gene therapy; anticancer;
KW   antimetastatic; vaccine; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965928-A2.
XX
PD   23-DEC-1999.
XX
PF   18-JUN-1999;   99WO-US13647.
XX
PR   19-JUN-1998;   98US-0089853.
PR   19-JUN-1998;   98US-0089997.
PR   19-JUN-1998;   98US-0090039.
PR   19-JUN-1998;   98US-0090040.
PR   19-JUN-1998;   98US-0090041.
XX
PA   (GENZ ) GENZYME CORP.
PA   (ROBE/) ROBERTS B L.
PA   (SHAN/) SHANKARA S.
XX
PI   Roberts BL,  Shankara S;
XX
DR   WPI; 2000-106079/09.
XX
PT   Isolated polynucleotides differentially expressed between metastatic
PT   and non-metastatic breast cancer cells, useful for diagnosis,
PT   prevention and treatment of cancer -
XX
PS   Claim 1; Page 114; 219pp; English.
XX
CC   AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC   transcripts that are preferentially transcribed in the metastatic breast
CC   tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC   AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC   that are preferentially transcribed in the primary or non-metastatic
CC   breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC   cells). These transcripts can be used for diagnosis, prognosis,
CC   monitoring and treatment of breast cancer, particularly where metastatic.
CC   Diagnosis is by standard immunoassays or hybridisation/amplification
CC   reactions. Compounds that modulate expression of the transcripts are
CC   potentially useful for treatment of (metastatic) breast cancer, while
CC   promoters from the transcripts are used to direct expression, in selected
CC   cell types, of e.g. therapeutic genes (also ribozymes or antisense
```

CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 3 A; 1 C; 5 G; 1 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 7 CCGTTC 2

RESULT 14
AAZ83972
ID AAZ83972 standard; DNA; 10 BP.
XX
AC AAZ83972;
XX
DT 07-APR-2000 (first entry)
XX
DE Metastatic breast tumour cell downregulated transcript tag #3206.
XX
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW non-metastatic breast tumour tissue; gene therapy; anticancer;
KW antimetastatic; vaccine; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9965928-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13647.
XX
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106079/09.
XX
PT Isolated polynucleotides differentially expressed between metastatic
PT and non-metastatic breast cancer cells, useful for diagnosis,
PT prevention and treatment of cancer -
XX
PS Claim 1; Page 145; 219pp; English.
XX
CC AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC transcripts that are preferentially transcribed in the metastatic breast
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC that are preferentially transcribed in the primary or non-metastatic
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC cells). These transcripts can be used for diagnosis, prognosis,
CC monitoring and treatment of breast cancer, particularly where metastatic.
CC Diagnosis is by standard immunoassays or hybridisation/amplification
CC reactions. Compounds that modulate expression of the transcripts are
CC potentially useful for treatment of (metastatic) breast cancer, while
CC promoters from the transcripts are used to direct expression, in selected
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 1 A; 3 C; 2 G; 4 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 2 ccgttc 7

RESULT 15
AAZ84187
ID AAZ84187 standard; DNA; 10 BP.
XX
AC AAZ84187;
XX
DT 07-APR-2000 (first entry)
XX
DE Metastatic breast tumour cell downregulated transcript tag #3421.
XX
KW Human; metastatic breast tumour tissue; breast cancer; tag; primer;
KW non-metastatic breast tumour tissue; gene therapy; anticancer;
KW antimetastatic; vaccine; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9965928-A2.
XX
PD 23-DEC-1999.
XX
PF 18-JUN-1999; 99WO-US13647.
XX
PR 19-JUN-1998; 98US-0089853.
PR 19-JUN-1998; 98US-0089997.
PR 19-JUN-1998; 98US-0090039.
PR 19-JUN-1998; 98US-0090040.
PR 19-JUN-1998; 98US-0090041.
XX
PA (GENZ) GENZYME CORP.
PA (ROBE/) ROBERTS B L.
PA (SHAN/) SHANKARA S.
XX
PI Roberts BL, Shankara S;
XX
DR WPI; 2000-106079/09.
XX
PT Isolated polynucleotides differentially expressed between metastatic
PT and non-metastatic breast cancer cells, useful for diagnosis,
PT prevention and treatment of cancer -
XX
PS Claim 1; Page 150; 219pp; English.
XX
CC AAZ80767 to AAZ83941 represent tags corresponding to distinct
CC transcripts that are preferentially transcribed in the metastatic breast
CC tumour tissue (i.e. are upregulated in metastatic breast tumour cells).
CC AAZ83942 to AAZ86677 represent tags corresponding to distinct transcripts
CC that are preferentially transcribed in the primary or non-metastatic
CC breast tumour tissue (i.e. are downregulated in metastatic breast tumour
CC cells). These transcripts can be used for diagnosis, prognosis,

CC monitoring and treatment of breast cancer, particularly where metastatic.
CC Diagnosis is by standard immunoassays or hybridisation/amplification
CC reactions. Compounds that modulate expression of the transcripts are
CC potentially useful for treatment of (metastatic) breast cancer, while
CC promoters from the transcripts are used to direct expression, in selected
CC cell types, of e.g. therapeutic genes (also ribozymes or antisense
CC sequences), particularly an antigen-encoding sequence for use in gene or
CC cell-based vaccines. Polypeptides encoded by the transcripts are also
CC useful in vaccines; for diagnosing breast cancer and for raising
CC specific antibodies (Ab). Ab are used to detect the polypeptides or as
CC therapeutic agents. Host cells that produce the polypeptides can be used
CC to expand and isolate populations of educated, antigen-specific immune
CC effecter cells, e.g. cytotoxic T lymphocytes, and these used for
CC adoptive immunotherapy.
XX
SQ Sequence 10 BP; 0 A; 5 C; 3 G; 2 T; 0 other;

Query Match 100.0%; Score 6; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 2 ccgttc 7

Search completed: July 30, 2002, 00:01:19
Job time: 4930 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:22:40 ; Search time 2542.47 Seconds
(without alignments)
31.852 Million cell updates/sec

Title: US-09-530-663B-16
Perfect score: 6
Sequence: 1 ccgttc 6

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	6	100.0	19	9	AA911671	AA911671 ol49f08.s
c 2	6	100.0	21	12	AZ347845	AZ347845 1M0084L06
c 3	6	100.0	22	12	AZ811866	AZ811866 2M0078M08
c 4	6	100.0	23	12	AZ653869	AZ653869 1M0527D14
5	6	100.0	24	12	AZ316663	AZ316663 1M0034G22
c 6	6	100.0	24	12	AZ436588	AZ436588 1M0224H20
c 7	6	100.0	24	12	AZ446206	AZ446206 1M0242I06
c 8	6	100.0	24	12	AZ662500	AZ662500 1M0541G07
9	6	100.0	24	12	AZ789936	AZ789936 2M0038L17
10	6	100.0	24	12	TA143H10Q	AL467041 T. brucei
c 11	6	100.0	25	9	AI199669	AI199669 qi60a02.x
c 12	6	100.0	25	12	AZ442576	AZ442576 1M0236K13
c 13	6	100.0	25	12	AZ621312	AZ621312 1M0454P19
c 14	6	100.0	26	12	AZ372925	AZ372925 1M0125D04
15	6	100.0	26	12	AZ819947	AZ819947 2M0091E20
16	6	100.0	26	12	TA347D12P	AL493847 T. brucei
17	6	100.0	27	10	BG927944	BG927944 HNC45-1-F
c 18	6	100.0	27	12	AZ488404	AZ488404 1M0318I11
19	6	100.0	27	12	AZ494628	AZ494628 1M0330E06
c 20	6	100.0	27	12	AZ809974	AZ809974 2M0074C18
21	6	100.0	27	12	TA6F09Q	AL451746 T. brucei
22	6	100.0	28	9	AI748505	AI748505 sb53h08.y
23	6	100.0	28	12	AZ780363	AZ780363 2M0017I11
24	6	100.0	28	12	AZ830168	AZ830168 2M0109E13
c 25	6	100.0	28	12	TA116E03P	AL462526 T. brucei
c 26	6	100.0	29	12	AZ595520	AZ595520 1M0408M09
27	6	100.0	29	12	TA119F03Q	AL463260 T. brucei
c 28	6	100.0	30	10	BG719541	BG719541 602690091
c 29	6	100.0	30	10	BG719681	BG719681 602689891
c 30	6	100.0	31	9	AI118841	AI118841 uc14d10.r
c 31	6	100.0	31	9	AI174159	AI174159 vz84e05.r
c 32	6	100.0	31	9	AI736496	AI736496 sb29d11.y
c 33	6	100.0	31	10	N94283	N94283 za26f01.r1
c 34	6	100.0	31	12	TA106A06P	AL459090 T. brucei
35	6	100.0	32	10	BF168323	BF168323 601774306
c 36	6	100.0	32	12	AZ642287	AZ642287 1M0505D16
37	6	100.0	32	12	TA121A02P	AL462540 T. brucei
38	6	100.0	32	12	TA12F01Q	AL451365 T. brucei
c 39	6	100.0	33	12	AZ379585	AZ379585 1M0134P16
c 40	6	100.0	34	9	AI225227	AI225227 qx12b04.x
c 41	6	100.0	34	9	AI971896	AI971896 wv29g12.x
42	6	100.0	34	10	BI692262	BI692262 603342729
43	6	100.0	34	10	BJ055330	BJ055330 BJ055330
44	6	100.0	34	12	AQ050771	AQ050771 nbxb0001c
45	6	100.0	34	12	AZ438586	AZ438586 1M0228E15

ALIGNMENTS

RESULT 1
AA911671/c
LOCUS AA911671 19 bp mRNA linear EST 10-JUN-1998
DEFINITION ol49f08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1526823 3' similar to TR:Q18444 Q18444 COSMID C34D4. ;contains MSR1.b2 MSR1 repetitive element ;, mRNA sequence.
ACCESSION AA911671
VERSION AA911671.1 GI:3051035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 19)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 682 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .19
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1526823"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 3 a 2 c 11 g 3 t
ORIGIN

Query Match 100.0%; Score 6; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.9e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 11 CCGTTC 6

RESULT 2
AZ347845/c
LOCUS AZ347845 21 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0084L06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0084L06 F, DNA sequence.
ACCESSION AZ347845
VERSION AZ347845.1 GI:10427082
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 21)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0084 row: L column: 06
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 21.
FEATURES Location/Qualifiers
source 1. .21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0084L06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 11 a 4 c 6 g 0 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 15 CCGTTC 10

RESULT 3
AZ811866/c
LOCUS AZ811866 22 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0078M08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0078M08 F, DNA sequence.
ACCESSION AZ811866
VERSION AZ811866.1 GI:12980548
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 22)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0078 row: M column: 08
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 22.
FEATURES Location/Qualifiers
source 1. .22
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0078M08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 8 a 5 c 7 g 2 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 16 CCGTTC 11

RESULT 4
AZ653869/c
LOCUS AZ653869 23 bp DNA linear GSS 14-DEC-2000
DEFINITION 1M0527D14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0527D14 R, DNA sequence.
ACCESSION AZ653869
VERSION AZ653869.1 GI:11791015
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 23)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0527 row: D column: 14
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 23.
FEATURES Location/Qualifiers
source 1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0527D14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 4 a 1 c 13 g 5 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 23;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 11 CCGTTC 6

RESULT 5
AZ316663
LOCUS AZ316663 24 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0034G22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0034G22 R, DNA sequence.
ACCESSION AZ316663
VERSION AZ316663.1 GI:10364703
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0034 row: G column: 22
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0034G22"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 5 a 9 c 4 g 6 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 9 CCGTTC 14

RESULT 6
AZ436588/c
LOCUS AZ436588 24 bp DNA linear GSS 03-OCT-2000
DEFINITION 1M0224H20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0224H20 F, DNA sequence.
ACCESSION AZ436588
VERSION AZ436588.1 GI:10560601
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0224 row: H column: 20
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0224H20"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 3 a 5 c 13 g 3 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 12 CCGTTC 7

RESULT 7
AZ446206/c
LOCUS AZ446206 24 bp DNA linear GSS 04-OCT-2000
DEFINITION 1M0242I06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0242I06 R, DNA sequence.
ACCESSION AZ446206
VERSION AZ446206.1 GI:10596787
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0242 row: I column: 06
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0242I06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 4 a 6 c 10 g 4 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 20 CCGTTC 15

RESULT 8
AZ662500/c
LOCUS AZ662500 24 bp DNA linear GSS 14-DEC-2000
DEFINITION 1M0541G07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0541G07 R, DNA sequence.
ACCESSION AZ662500
VERSION AZ662500.1 GI:11799646
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0541 row: G column: 07
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0541G07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 8 a 3 c 7 g 6 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 22 CCGTTC 17

RESULT 9
AZ789936
LOCUS AZ789936 24 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0038L17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0038L17 F, DNA sequence.
ACCESSION AZ789936
VERSION AZ789936.1 GI:12931470
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0038 row: L column: 17
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 24.
FEATURES Location/Qualifiers
source 1. .24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0038L17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 4 a 10 c 5 g 5 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 5 CCGTTC 10

RESULT 10
TA143H10Q
LOCUS TA143H10Q 24 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 143h10, reverse sequence, genomic survey sequence.
ACCESSION AL467041
VERSION AL467041.1 GI:11836396
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 24)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES Location/Qualifiers
source 1. .24
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="143h10"
BASE COUNT 2 a 3 c 9 g 10 t
ORIGIN

Query Match 100.0%; Score 6; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 2 CCGTTC 7

RESULT 11
AI199669/c
LOCUS AI199669 25 bp mRNA linear EST 02-DEC-1998
DEFINITION qi60a02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1860842 3' similar to SW:PAXI_HUMAN P49023 PAXILLIN. [2] TR:O14970 ;, mRNA sequence.
ACCESSION AI199669
VERSION AI199669.1 GI:3752275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 25)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 654 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .25
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1860842"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 5 a 7 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 6; DB 9; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttc 6
||||||
Db 18 CCGTTC 13

RESULT 12
AZ442576/c
LOCUS AZ442576 25 bp DNA linear GSS 04-OCT-2000

```
DEFINITION  1M0236K13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0236K13 R, DNA sequence.
ACCESSION   AZ442576
VERSION     AZ442576.1  GI:10589722
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0236  row: K  column: 13
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0236K13"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        9 a      6 c      6 g      4 t
ORIGIN


  Query Match             100.0%;  Score 6;  DB 12;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db       9 CCGTTC 4


RESULT  13
AZ621312/c
```

```
LOCUS       AZ621312                25 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0454P19F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0454P19 F, DNA sequence.
ACCESSION   AZ621312
VERSION     AZ621312.1  GI:11743502
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0454  row: P  column: 19
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0454P19"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        4 a      2 c     15 g      4 t
ORIGIN


  Query Match             100.0%;  Score 6;  DB 12;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 6.1e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db      15 CCGTTC 10


RESULT  14
```

```
AZ372925/c
LOCUS       AZ372925                 26 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0125D04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0125D04 F, DNA sequence.
ACCESSION   AZ372925
VERSION     AZ372925.1  GI:10486625
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0125  row: D  column: 04
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 26.
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0125D04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        8 a      4 c      9 g      5 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 12;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 6.2e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db      16 CCGTTC 11
```

```
RESULT  15
AZ819947
LOCUS       AZ819947                 26 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0091E20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0091E20 R, DNA sequence.
ACCESSION   AZ819947
VERSION     AZ819947.1  GI:12989855
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 26)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0091  row: E  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 26.
FEATURES             Location/Qualifiers
     source          1. .26
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0091E20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        3 a      9 c      4 g      10 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 12;  Length 26;
  Best Local Similarity  100.0%;  Pred. No. 6.2e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db      13 CCGTTC 18
```

Search completed: July 29, 2002, 23:22:44
Job time: 6820 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:56:13 ; Search time 65.09 Seconds
(without alignments)
22.643 Million cell updates/sec

Title: US-09-530-663B-16
Perfect score: 6
Sequence: 1 ccgttc 6

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	6	100.0	9	2	US-08-642-045B-2	Sequence 2, Appli
c 2	6	100.0	9	3	US-08-852-268-4	Sequence 4, Appli
c 3	6	100.0	10	2	US-08-545-253A-8	Sequence 8, Appli
c 4	6	100.0	10	3	US-08-719-337-8	Sequence 8, Appli
c 5	6	100.0	10	4	US-08-878-835A-7	Sequence 7, Appli
6	6	100.0	12	1	US-07-974-447-10	Sequence 10, Appl
7	6	100.0	12	1	US-08-149-199-10	Sequence 10, Appl
c 8	6	100.0	12	1	US-08-411-727-5	Sequence 5, Appli
9	6	100.0	12	1	US-08-411-727-6	Sequence 6, Appli
10	6	100.0	12	2	US-08-858-767-6	Sequence 6, Appli
11	6	100.0	12	2	US-08-858-767-8	Sequence 8, Appli
12	6	100.0	12	2	US-08-863-028-6	Sequence 6, Appli
13	6	100.0	12	2	US-08-863-028-8	Sequence 8, Appli
14	6	100.0	12	3	US-09-115-061-10	Sequence 10, Appl
15	6	100.0	12	4	US-09-261-079-10	Sequence 10, Appl
c 16	6	100.0	14	1	US-08-424-921-5	Sequence 5, Appli
17	6	100.0	14	1	US-08-424-921-6	Sequence 6, Appli
18	6	100.0	14	1	US-08-663-769-11	Sequence 11, Appl
19	6	100.0	14	3	US-08-193-627-2	Sequence 2, Appli
20	6	100.0	14	3	US-08-985-162-1835	Sequence 1835, Ap
21	6	100.0	14	3	US-08-985-162-1836	Sequence 1836, Ap
c 22	6	100.0	15	1	US-08-182-968A-26	Sequence 26, Appl
c 23	6	100.0	15	1	US-08-182-968A-27	Sequence 27, Appl
c 24	6	100.0	15	2	US-08-774-306A-26	Sequence 26, Appl
c 25	6	100.0	15	2	US-08-774-306A-27	Sequence 27, Appl
c 26	6	100.0	15	3	US-09-064-156A-26	Sequence 26, Appl
c 27	6	100.0	15	3	US-09-064-156A-27	Sequence 27, Appl
c 28	6	100.0	15	3	US-08-461-366A-4	Sequence 4, Appli
c 29	6	100.0	15	3	US-08-481-341-14	Sequence 14, Appl
30	6	100.0	15	4	US-09-402-764B-11	Sequence 11, Appl
c 31	6	100.0	15	4	US-08-464-514-6	Sequence 6, Appli
c 32	6	100.0	15	4	US-08-486-403-6	Sequence 6, Appli
c 33	6	100.0	15	4	US-09-078-954-17	Sequence 17, Appl
c 34	6	100.0	15	4	US-09-054-832-26	Sequence 26, Appl
c 35	6	100.0	15	5	PCT-US94-00265-14	Sequence 14, Appl
c 36	6	100.0	15	6	5182195-49	Patent No. 5182195
37	6	100.0	16	1	US-08-513-841-14	Sequence 14, Appl
38	6	100.0	16	2	US-08-696-834-15	Sequence 15, Appl
39	6	100.0	16	2	US-08-942-673-14	Sequence 14, Appl
40	6	100.0	16	4	US-09-118-317-14	Sequence 14, Appl
c 41	6	100.0	16	4	US-09-134-607A-3	Sequence 3, Appli
c 42	6	100.0	16	4	US-08-679-645-521	Sequence 521, App
c 43	6	100.0	17	1	US-07-879-647A-43	Sequence 43, Appl
c 44	6	100.0	17	1	US-07-879-584A-43	Sequence 43, Appl
c 45	6	100.0	17	1	US-07-879-470A-43	Sequence 43, Appl

ALIGNMENTS

```
RESULT 1
US-08-642-045B-2/c
; Sequence 2, Application US/08642045B
; Patent No. 5851804
;  GENERAL INFORMATION:
;    APPLICANT:  Snyder, Linda A.
;    APPLICANT:  Satishchandran, C.
;    TITLE OF INVENTION:  CHIMERIC KANAMYCIN RESISTANCE GENE
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5851804ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS
;      SOFTWARE:  WordPerfect 6.0/6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/642,045B
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  APOL-0262
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  DNA
US-08-642-045B-2


  Query Match           100.0%;  Score 6;  DB 2;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 2.5e+07;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db       8 CCGTTC 3
```

```
RESULT   2
US-08-852-268-4/c
; Sequence 4, Application US/08852268
; Patent No. 6143527
;  GENERAL INFORMATION:
;    APPLICANT:  Pachuk, Catherine J.
;    APPLICANT:  Samuel, Manoj
;    APPLICANT:  Zurawski, John A.
;    APPLICANT:  Satishchandran, C.
;    TITLE OF INVENTION:  CHAIN REACTION CLONING
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 6143527ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/852,268
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/642,045
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  APOL-0265
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  DNA
US-08-852-268-4


  Query Match           100.0%;  Score 6;  DB 3;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 2.5e+07;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db       8 CCGTTC 3


RESULT   3
US-08-545-253A-8/c
; Sequence 8, Application US/08545253A
; Patent No. 5908978
;  GENERAL INFORMATION:
;    APPLICANT:  O'Malley, David M.
;    APPLICANT:  Sederoff, Ronald R.
;    APPLICANT:  Grattapaglia, Dario
;    APPLICANT:  Henry V. Amerson
;    APPLICANT:  Phillip Wilcox
;    APPLICANT:  E. George Kuhlman
;    TITLE OF INVENTION:  METHODS FOR WITHIN FAMILY
;    TITLE OF INVENTION:  SELECTION IN
;    TITLE OF INVENTION:  WOODY PERENNIALS USING GENETIC MARKERS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kenneth D. Sibley
;      STREET:  Post Office Drawer 34009
;      CITY:  Charlotte
;      STATE:  No. 5908978th Carolina
;      COUNTRY:  U.S.A.
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,253A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5051-281
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 881-3140
;      TELEFAX:  (919) 881-3175
;      TELEX:  575102
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-545-253A-8


  Query Match           100.0%;  Score 6;  DB 2;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db       6 CCGTTC 1


RESULT   4
US-08-719-337-8/c
; Sequence 8, Application US/08719337
; Patent No. 6054634
;  GENERAL INFORMATION:
;    APPLICANT:  O'Malley, David M.
;    APPLICANT:  Sederoff, Ronald R.
;    APPLICANT:  Grattapaglia, Dario
;    TITLE OF INVENTION:  METHODS FOR WITHIN FAMILY SELECTION IN
;    TITLE OF INVENTION:  WOODY PERENNIALS USING GENETIC MARKERS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kenneth D. Sibley
;      STREET:  Post Office Drawer 34009
;      CITY:  Charlotte
;      STATE:  No. 6054634th Carolina
;      COUNTRY:  U.S.A.
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,337
;      FILING DATE:  25-SEP-1996
;      CLASSIFICATION:  047
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/184,567
;      FILING DATE:  21-JAN-1994
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5051-247
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 881-3140
;      TELEFAX:  (919) 881-3175
;      TELEX:  575102
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-719-337-8


  Query Match            100.0%;  Score 6;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      6 CCGTTC 1


RESULT   5
US-08-878-835A-7/c
; Sequence 7, Application US/08878835A
; Patent No. 6337071
;  GENERAL INFORMATION:
;    APPLICANT:  William Mitchell Molyneux
;    TITLE OF INVENTION:  Mosquito and/or Flea Control
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  D. Peter Hochberg Co., L.P.A.
;      STREET:  The Baker Building - Sixth Floor 1940 East 6th Street
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MByte storage
;      COMPUTER:  IBM Compatible w/ Pentium Processor
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Microsoft Word 97
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,835A
;      FILING DATE:  June 19, 1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PO 0605
;      FILING DATE:  21 June 1996
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
US-08-878-835A-7


  Query Match            100.0%;  Score 6;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      6 CCGTTC 1


RESULT   6
US-07-974-447-10
; Sequence 10, Application US/07974447
; Patent No. 5436142
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H
;    APPLICANT:  Lisitsyn, Nikolai
;    TITLE OF INVENTION:  A REPRESENTATIONAL APPROACH TO GENOMIC
;    TITLE OF INVENTION:  ANALYSIS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/974,447
;      FILING DATE:  12-NOV-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A-57438/BIR CSHL-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-07-974-447-10


  Query Match            100.0%;  Score 6;  DB 1;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT   7
US-08-149-199-10
; Sequence 10, Application US/08149199
; Patent No. 5501964
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H
;    APPLICANT:  Lisitsyn, Nikolai
;    TITLE OF INVENTION:  A REPRESENTATIONAL APPROACH TO GENOMIC
;    TITLE OF INVENTION:  ANALYSIS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  941114187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PCDOS/MSDOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,199
;      FILING DATE:  9-No. 5501964-93
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A57438/BIR CSHL002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 7811989
;      TELEFAX:  (415) 3983249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-149-199-10


  Query Match             100.0%;  Score 6;  DB 1;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db       4 CCGTTC 9


RESULT   8
US-08-411-727-5/c
; Sequence 5, Application US/08411727
; Patent No. 5705161
; Patent No. 5705161 5683703
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DER LEY, Peter Andre
;    APPLICANT:  POOLMAN, Jan Theunis
;    APPLICANT:  HOOGERHOUT, Peter
;    TITLE OF INVENTION:  IMMUNOGENIC MENINGOCOCCAL LPS AND OTHER
;    TITLE OF INVENTION:  MEMBRANE VESICLES AND VACCINE THEREFROM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  YOUNG & THOMPSON
;      STREET:  745 South 23rd Street, Suite 200
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,727
;      FILING DATE:  01-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  NL 9201716
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/NL93/00163
;      FILING DATE:  30-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATCH, Andrew J.
;      REGISTRATION NUMBER:  32925
;      REFERENCE/DOCKET NUMBER:  BO 38275
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-521-2297
;      TELEFAX:  703-685-0573
;      TELEX:  248425 EMBON
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-411-727-5


  Query Match             100.0%;  Score 6;  DB 1;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccgttc 6
           ||||||
Db      11 CCGTTC 6


RESULT   9
US-08-411-727-6
; Sequence 6, Application US/08411727
; Patent No. 5705161
; Patent No. 5705161 5683703
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DER LEY, Peter Andre
;    APPLICANT:  POOLMAN, Jan Theunis
;    APPLICANT:  HOOGERHOUT, Peter
;    TITLE OF INVENTION:  IMMUNOGENIC MENINGOCOCCAL LPS AND OTHER
;    TITLE OF INVENTION:  MEMBRANE VESICLES AND VACCINE THEREFROM
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  YOUNG & THOMPSON
;      STREET:  745 South 23rd Street, Suite 200
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,727
;      FILING DATE:  01-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  NL 9201716
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/NL93/00163
;      FILING DATE:  30-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATCH, Andrew J.
;      REGISTRATION NUMBER:  32925
;      REFERENCE/DOCKET NUMBER:  BO 38275
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-521-2297
;      TELEFAX:  703-685-0573
;      TELEX:  248425 EMBON
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-411-727-6


  Query Match             100.0%;  Score 6;  DB 1;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
```

```
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      2 CCGTTC 7


RESULT  10
US-08-858-767-6
; Sequence 6, Application US/08858767
; Patent No. 5837468
;  GENERAL INFORMATION:
;    APPLICANT:  WANG, Xun
;    APPLICANT:  DUVICK, Jonathan P.
;    APPLICANT:  BRIGGS, Steven P.
;    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
;    TITLE OF INVENTION:  METHOD
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
;      FILING DATE:  19-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/481,687
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-858-767-6


  Query Match             100.0%;  Score 6;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  11
US-08-858-767-8
; Sequence 8, Application US/08858767
; Patent No. 5837468
;  GENERAL INFORMATION:
;    APPLICANT:  WANG, Xun
;    APPLICANT:  DUVICK, Jonathan P.
;    APPLICANT:  BRIGGS, Steven P.
;    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
;    TITLE OF INVENTION:  METHOD
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
;      FILING DATE:  19-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/481,687
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-858-767-8


  Query Match             100.0%;  Score 6;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  12
US-08-863-028-6
; Sequence 6, Application US/08863028
; Patent No. 5853991
;  GENERAL INFORMATION:
;    APPLICANT:  WANG, Xun
;    APPLICANT:  DUVICK, Jonathan P.
;    APPLICANT:  BRIGGS, Steven P.
;    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
;    TITLE OF INVENTION:  METHOD
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/863,028
;      FILING DATE:
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
;      FILING DATE:  19-MAY-1997
;      APPLICATION NUMBER:  US 08/481,687
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-863-028-6


  Query Match           100.0%;  Score 6;  DB 2;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  13
US-08-863-028-8
; Sequence 8, Application US/08863028
; Patent No. 5853991
;  GENERAL INFORMATION:
;    APPLICANT:  WANG, Xun
;    APPLICANT:  DUVICK, Jonathan P.
;    APPLICANT:  BRIGGS, Steven P.
;    TITLE OF INVENTION:  PCR-BASED CDNA SUBTRACTIVE CLONING
;    TITLE OF INVENTION:  METHOD
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/863,028
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,767
;      FILING DATE:  19-MAY-1997
;      APPLICATION NUMBER:  US 08/481,687
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/325/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-863-028-8


  Query Match           100.0%;  Score 6;  DB 2;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  14
US-09-115-061-10
; Sequence 10, Application US/09115061A
; Patent No. 6159713
; GENERAL INFORMATION:
;  APPLICANT: Wigler, Michael
;  APPLICANT: Lisitsyn, Nikolai
;  TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO DNA ANALYSIS
;  FILE REFERENCE: CSHL.002.03US
;  CURRENT APPLICATION NUMBER: US/09/115,061A
;  CURRENT FILING DATE: 1998-07-14
;  EARLIER APPLICATION NUMBER: 08/478,242
;  EARLIER FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 08/149,199
;  EARLIER FILING DATE: 1993-11-09
;  EARLIER APPLICATION NUMBER: 07/974,447
;  EARLIER FILING DATE: 1992-11-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 12
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: OLIGONUCLEOTIDE
US-09-115-061-10


  Query Match           100.0%;  Score 6;  DB 3;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+04;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 ccgttc 9


RESULT  15
US-09-261-079-10
; Sequence 10, Application US/09261079
; Patent No. 6277606
; GENERAL INFORMATION:
;  APPLICANT: Wigler, Michael
;  APPLICANT: Lisitsyn, Nikolai
;  TITLE OF INVENTION: A REPRESENTATIONAL APPROACH TO DNA ANALYSIS
;  FILE REFERENCE: CSHL.002.04US
;  CURRENT APPLICATION NUMBER: US/09/261,079
;  CURRENT FILING DATE: 1999-03-02
;  EARLIER APPLICATION NUMBER: 08/478,242
;  EARLIER FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 07/974,447
;  EARLIER FILING DATE: 1992-11-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 12
```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: OLIGONUCLEOTIDE
US-09-261-079-10

Query Match 100.0%; Score 6; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ccgttc 6
          ||||||
Db      4 ccgttc 9
```

Search completed: July 29, 2002, 23:56:14
Job time: 4755 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 29, 2002, 23:55:03 ; Search time 1921.77 Seconds
(without alignments)
65.335 Million cell updates/sec

Title: US-09-530-663B-16
Perfect score: 6
Sequence: 1 ccgttc 6

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	6	100.0	10	6	AR070974	AR070974 Sequence
2	6	100.0	10	6	AX043779	AX043779 Sequence
c 3	6	100.0	10	6	AX152405	AX152405 Sequence
c 4	6	100.0	10	6	AX152684	AX152684 Sequence
5	6	100.0	10	6	AX301586	AX301586 Sequence
c 6	6	100.0	12	6	A38146	A38146 Sequence 2
7	6	100.0	12	6	AR055101	AR055101 Sequence
8	6	100.0	12	6	AR055103	AR055103 Sequence
9	6	100.0	12	6	AR068442	AR068442 Sequence
10	6	100.0	12	6	AR068444	AR068444 Sequence
11	6	100.0	12	6	AR121283	AR121283 Sequence
12	6	100.0	12	6	I13336	I13336 Sequence 10
13	6	100.0	12	6	I19029	I19029 Sequence 10
14	6	100.0	14	6	BD000935	BD000935 Method an
15	6	100.0	14	6	BD001364	BD001364 Method an
c 16	6	100.0	14	6	I24766	I24766 Sequence 5
17	6	100.0	14	6	I24767	I24767 Sequence 6
18	6	100.0	14	6	I38426	I38426 Sequence 11
c 19	6	100.0	15	6	A01751	A01751 DNA fragmen
c 20	6	100.0	15	6	A11090	A11090 Oligonucleo
21	6	100.0	15	6	A84556	A84556 Sequence 11
c 22	6	100.0	15	6	A89485	A89485 Sequence 16
c 23	6	100.0	15	6	AR033260	AR033260 Sequence
c 24	6	100.0	15	6	AR033261	AR033261 Sequence
c 25	6	100.0	15	6	AR113082	AR113082 Sequence
c 26	6	100.0	15	6	AR113083	AR113083 Sequence
c 27	6	100.0	15	6	AR116412	AR116412 Sequence
c 28	6	100.0	15	6	AR121617	AR121617 Sequence
29	6	100.0	15	6	AR144754	AR144754 Sequence
c 30	6	100.0	15	6	AR166544	AR166544 Sequence
c 31	6	100.0	15	6	AR167451	AR167451 Sequence
c 32	6	100.0	15	6	AR176695	AR176695 Sequence
c 33	6	100.0	15	6	AX108761	AX108761 Sequence
34	6	100.0	15	6	AX196235	AX196235 Sequence
c 35	6	100.0	15	6	I57489	I57489 Sequence 26
c 36	6	100.0	15	6	I57490	I57490 Sequence 27
37	6	100.0	16	6	AR008356	AR008356 Sequence
38	6	100.0	16	6	AR030665	AR030665 Sequence
39	6	100.0	16	6	AR053771	AR053771 Sequence
40	6	100.0	16	6	AR137835	AR137835 Sequence
c 41	6	100.0	16	6	AX074441	AX074441 Sequence
c 42	6	100.0	16	6	AX138221	AX138221 Sequence
c 43	6	100.0	16	6	AX255602	AX255602 Sequence
c 44	6	100.0	16	6	AX255636	AX255636 Sequence
c 45	6	100.0	16	6	AX317642	AX317642 Sequence

ALIGNMENTS

RESULT 1
AR070974/c
LOCUS AR070974 10 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 8 from patent US 5908978.
ACCESSION AR070974
VERSION AR070974.1 GI:7221862
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 10)
AUTHORS Amerson,H.V., Wilcox,P., Sederoff,R.R., Kuhlman,E.George, O'Malley,D.M. and Grattapaglia,D.
TITLE Methods for within family selection of disease resistance in woody perennials using genetic markers
JOURNAL Patent: US 5908978-A 8 01-JUN-1999;
FEATURES Location/Qualifiers
source 1. .10
/organism="unknown"
BASE COUNT 3 a 3 c 3 g 1 t
ORIGIN

```
  Query Match           100.0%;  Score 6;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      6 CCGTTC 1


RESULT   2
AX043779
LOCUS       AX043779                 10 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 11 from Patent WO0065073.
ACCESSION   AX043779
VERSION     AX043779.1  GI:11342383
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Martens,S. and Forkmann,G.
  TITLE     Genetic sequence which codes for the flavon synthase ii enzyme and
            use of the same
  JOURNAL   Patent: WO 0065073-A 11 02-NOV-2000;
            Martens, Stefan (DE) ; Forkmann, Gert (DE)
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="synthetisches Oligonukleotid"
BASE COUNT        0 a      4 c      4 g      2 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      3 CCGTTC 8


RESULT   3
AX152405/c
LOCUS       AX152405                 10 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 320 from Patent WO0138577.
ACCESSION   AX152405
VERSION     AX152405.1  GI:14534056
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Velculescu,V.E., Vogelstein,B. and Kinzler,K.W.
  TITLE     Human transcriptomes
  JOURNAL   Patent: WO 0138577-A 320 31-MAY-2001;
            The Johns Hopkins University (US)
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        3 a      2 c      5 g      0 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      9 CCGTTC 4


RESULT   4
AX152684/c
LOCUS       AX152684                 10 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 599 from Patent WO0138577.
ACCESSION   AX152684
VERSION     AX152684.1  GI:14534335
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10)
  AUTHORS   Velculescu,V.E., Vogelstein,B. and Kinzler,K.W.
  TITLE     Human transcriptomes
  JOURNAL   Patent: WO 0138577-A 599 31-MAY-2001;
            The Johns Hopkins University (US)
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        3 a      1 c      6 g      0 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      8 CCGTTC 3


RESULT   5
AX301586
LOCUS       AX301586                 10 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 300 from Patent WO0185941.
ACCESSION   AX301586
VERSION     AX301586.1  GI:17382669
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Versteeg,R. and Caron,H.N.
  TITLE     Myc targets
  JOURNAL   Patent: WO 0185941-A 300 15-NOV-2001;
            Academisch Ziekenhuis bij de Universiteit van Amsterdam (NL)
FEATURES             Location/Qualifiers
     source          1. .10
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        1 a      3 c      2 g      4 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 10;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      2 CCGTTC 7


RESULT   6
A38146/c
LOCUS       A38146                   12 bp    DNA     linear   PAT 05-MAR-1997
DEFINITION  Sequence 2 from Patent WO9408021.
```

```
ACCESSION   A38146
VERSION     A38146.1  GI:2294752
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Van,D.L., Poolman,J.T. and Hoogerhout,P.
  TITLE     IMMUNOGENIC MENINGOCOCCAL LPS AND OUTER MEMBRANE VESICLES AND
            VACCINE THEREFROM
  JOURNAL   Patent: WO 9408021-A 2 14-APR-1994;
            NEDERLANDEN STAAT (NL)
COMMENT     Other publication AU 4835193 940426
            Other publication NO 951181 950601
            Other publication FI 951535 950601
            Other publication NL 9201716 940502
            Other publication JP 8501940T 960305.
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /clone="PJBOO7"
BASE COUNT        3 a      3 c      5 g      1 t
ORIGIN


  Query Match            100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db     11 CCGTTC 6


RESULT   7
AR055101
LOCUS       AR055101                  12 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 6 from patent US 5837468.
ACCESSION   AR055101
VERSION     AR055101.1  GI:5980678
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wang,X., Duvick,J.P. and Briggs,S.P.
  TITLE     PCR-based cDNA substractive cloning method
  JOURNAL   Patent: US 5837468-A 6 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match            100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT   8
AR055103
LOCUS       AR055103                  12 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 8 from patent US 5837468.
ACCESSION   AR055103
VERSION     AR055103.1  GI:5980680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wang,X., Duvick,J.P. and Briggs,S.P.
  TITLE     PCR-based cDNA substractive cloning method
  JOURNAL   Patent: US 5837468-A 8 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match            100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT   9
AR068442
LOCUS       AR068442                  12 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 6 from patent US 5853991.
ACCESSION   AR068442
VERSION     AR068442.1  GI:6000649
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wang,X., Duvick,J.P. and Briggs,S.P.
  TITLE     PCR-based cDNA subtractive cloning method
  JOURNAL   Patent: US 5853991-A 6 29-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match            100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  10
AR068444
LOCUS       AR068444                  12 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 8 from patent US 5853991.
ACCESSION   AR068444
VERSION     AR068444.1  GI:6000651
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wang,X., Duvick,J.P. and Briggs,S.P.
  TITLE     PCR-based cDNA subtractive cloning method
  JOURNAL   Patent: US 5853991-A 8 29-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN
```

```
  Query Match           100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  11
AR121283
LOCUS       AR121283                  12 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 10 from patent US 6159713.
ACCESSION   AR121283
VERSION     AR121283.1  GI:14104859
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wigler,M. and Lisitsyn,N.
  TITLE     Methods for producing probes capable of distinguishing DNA from
            related sources
  JOURNAL   Patent: US 6159713-A 10 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  12
I13336
LOCUS       I13336                    12 bp    DNA     linear   PAT 26-JUL-1995
DEFINITION  Sequence 10 from patent US 5436142.
ACCESSION   I13336
VERSION     I13336.1  GI:910677
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wigler,M. and Lisitsyn,N.
  TITLE     Methods for producing probes capable of distingushing variant
            genomic sequences
  JOURNAL   Patent: US 5436142-A 10 25-JUL-1995;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  13
I19029
LOCUS       I19029                    12 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 10 from patent US 5501964.
ACCESSION   I19029
VERSION     I19029.1  GI:1599384
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Wigler,M. and Lisitsyn,N.
  TITLE     Methods for producing probes capable of distinguishing DNA from
            related sources
  JOURNAL   Patent: US 5501964-A 10 26-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
BASE COUNT        2 a      3 c      3 g      4 t
ORIGIN


  Query Match           100.0%;  Score 6;  DB 6;  Length 12;
  Best Local Similarity  100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  14
BD000935
LOCUS       BD000935                  14 bp    RNA     linear   PAT 31-JAN-2002
DEFINITION  Method and reagent for inhibiting viral replication.
ACCESSION   BD000935
VERSION     BD000935.1  GI:18625494
KEYWORDS    JP 2000342285-A/95.
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Draper,K.G., Dadyktz,L.W., Macswigen,J.A., Maysejak,D.G.,
            Holesek,J.J. and Mamone,A.J.
  TITLE     Method and reagent for inhibiting viral replication
  JOURNAL   Patent: JP 2000342285-A 95 12-DEC-2000;
            RIBOZYME PHARMACEUTICALS INC
COMMENT     OS   Artificial Sequence
            PN   JP 2000342285-A/95
            PD   12-DEC-2000
            PF   01-MAY-2000 JP 2000132616
            PR   11-MAY-1992 US    07/882689,14-MAY-1992 US    07/882712 PR
            14-MAY-1992 US    07/882713,14-MAY-1992 US    07/882714 PR
            14-MAY-1992 US    07/882823,14-MAY-1992 US    07/882824 PR
            14-MAY-1992 US    07/882886,14-MAY-1992 US    07/882888 PR
            14-MAY-1992 US    07/882889,14-MAY-1992 US    07/882921 PR
            14-MAY-1992 US    07/882922,14-MAY-1992 US    07/883823 PR
            14-MAY-1992 US    07/883849,14-MAY-1992 US    07/884073 PR
            14-MAY-1992 US    07/884074,14-MAY-1992 US    07/884333 PR
            14-MAY-1992 US    07/884422,14-MAY-1992 US    07/884431 PR
            14-MAY-1992 US    07/884436,14-MAY-1992 US    07/884521 PR
            31-JUL-1992 US    07/923738,26-AUG-1992 US    07/935854 PR
            26-AUG-1992 US    07/936086,18-SEP-1992 US    07/948359 PR
            15-OCT-1992 US    07/963322,07-DEC-1992 US    07/987129 PR
            07-DEC-1992 US    07/987130,07-DEC-1992 US    07/987133 PI
            KENNETH G DRAPER,LEC W DADYKTZ,JAMES A MACSWIGEN, PI   DENNIS G
            MAYSEJAK,
            PI   JAMES J HOLESEK,ANTHONY J MAMONE
            PC   C12N15/09,C12N5/10,C12N7/00,C12N9/22//(C12N5/10,C12R1:91), PC
             C12N15/00,
            PC   C12N5/00,(C12N5/00,C12R1:91)
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .14
            FT                   /organism='Artificial Sequence'.
```

```
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT        1 a      7 c      3 g      3 t
ORIGIN


  Query Match             100.0%;  Score 6;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


RESULT  15
BD001364
LOCUS       BD001364                 14 bp    RNA     linear   PAT 31-JAN-2002
DEFINITION  Method and reagent for inhibiting viral replication.
ACCESSION   BD001364
VERSION     BD001364.1  GI:18625923
KEYWORDS    JP 2000342286-A/95.
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 14)
  AUTHORS   Draper,K.G., Dadyktz,L.W., Macswigen,J.A., Maysejak,D.G.,
            Holesek,J.J. and Mamone,A.J.
  TITLE     Method and reagent for inhibiting viral replication
  JOURNAL   Patent: JP 2000342286-A 95 12-DEC-2000;
            RIBOZYME PHARMACEUTICALS INC
COMMENT     OS   Artificial Sequence
            PN   JP 2000342286-A/95
            PD   12-DEC-2000
            PF   01-MAY-2000 JP 2000132651
            PR   11-MAY-1992 US    07/882689,14-MAY-1992 US    07/882712 PR
            14-MAY-1992 US    07/882713,14-MAY-1992 US    07/882714 PR
            14-MAY-1992 US    07/882823,14-MAY-1992 US    07/882824 PR
            14-MAY-1992 US    07/882886,14-MAY-1992 US    07/882888 PR
            14-MAY-1992 US    07/882889,14-MAY-1992 US    07/882921 PR
            14-MAY-1992 US    07/882922,14-MAY-1992 US    07/883823 PR
            14-MAY-1992 US    07/883849,14-MAY-1992 US    07/884073 PR
            14-MAY-1992 US    07/884074,14-MAY-1992 US    07/884333 PR
            14-MAY-1992 US    07/884422,14-MAY-1992 US    07/884431 PR
            14-MAY-1992 US    07/884436,14-MAY-1992 US    07/884521 PR
            31-JUL-1992 US    07/923738,26-AUG-1992 US    07/935854 PR
            26-AUG-1992 US    07/936086,18-SEP-1992 US    07/948359 PR
            15-OCT-1992 US    07/963322,07-DEC-1992 US    07/987129 PR
            07-DEC-1992 US    07/987130,07-DEC-1992 US    07/987133 PI
            KENNETH G DRAPER,LEC W DADYKTZ,JAMES A MACSWIGEN, PI   DENNIS G
            MAYSEJAK,
            PI   JAMES J HOLESEK,ANTHONY J MAMONE
            PC   C12N15/09,C12N5/10,C12N7/00//A61K38/43,A61K39/125,A61K39/13,
            PC   A61K39/135,
            PC   A61K39/145,A61K39/21,A61K39/23,A61K39/245,A61K39/29,A61K48/00,
            PC   A61P1/16,
            PC   A61P31/14,A61P31/16,A61P31/18,A61P31/22,A61P35/02,C12Q1/68, PC
              (C12N15/09,C12R1:93),C12N15/00,C12N5/00,A61K37/48,(C12N15/00, PC
             C12R1:93)
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .14
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .14
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT        1 a      7 c      3 g      3 t
ORIGIN
```

```
  Query Match             100.0%;  Score 6;  DB 6;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttc 6
          ||||||
Db      4 CCGTTC 9


Search completed: July 29, 2002, 23:55:06
Job time: 7022 sec
```